

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

BULANIK KÜMELEMeye DAYALI PARAMETRE TAHMİNİ

YÜKSEK LİSANS TEZİ

Ersegül GÖKTÜRK ŞAHİN

OCAK 2014

TRABZON

KARADENİZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

BULANIK KÜMELEMeye DAYALI PARAMETRE TAHMİNİ

Ersegül GÖKTÜRK ŞAHİN

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"YÜKSEK LİSANS (İSTATİSTİK)"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

Tezin Enstitüye Verildiği Tarih : 02.01.2014

Tezin Savunma Tarihi : 24.01.2014

Tez Danışmanı: Doç. Dr. Türkan ERBAY DALKILIÇ

Trabzon 2014

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalında
Ersegül GÖKTÜRK ŞAHİN tarafından hazırlanan

BULANIK KÜMELEMeye DAYALI PARAMETRE TAHMİNİ

başlıklı bu çalışma, Enstitü Yönetim Kurulunun 07 /01/ 2014 gün ve 1536 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.

Jüri Üyeleri:

Başkan : Prof. Dr. İhsan ÜNVER

Üye : Doç. Dr. Zafer KÜÇÜK

Üye : Doç. Dr. Türkan ERBAY DALKILIÇ

Prof. Dr. Sadettin Korkmaz

Enstitü Müdürü

ÖNSÖZ

Bu tez çalışmasında bulanık mantık ile ilgili temel kavramlar, bulanık kümeleme algoritmaları, parametre tahmini ve bulanık kümelemeye dayalı parametre tahmini üzerinde durulmuştur. Bulanık kümelemeye dayalı parametre tahminleri, bulanık kümeleme yöntemlerinden Bulanık C-Ortalamlar (FCM) ve Gustafson-Kessel (G-K) algoritmaları kullanılarak elde edilen üyelik derecelerinden faydalanarak yapılmıştır. Bu tahminlere ilişkin hata miktarları, En Küçük Kareler (EKK) yönteminden elde edilen parametrelere bağlı tahminlerin hata miktarları ile karşılaştırılmıştır.

Öncelikle çalışmalarımı yönlendiren, araştırmalarımın her aşamasında bilgi, öneri ve yardımlarını esirgemeyen, yoğun çalışma temposunda bana zaman ayıran danışman hocam sayın Doç. Dr. Türkan ERBAY DALKILIÇ'a gönülden teşekkürü bir borç bilirim.

Yüksek lisans eğitimi yapma fırsatı bulduğum İstatistik ve Bilgisayar Bilimleri Bölümündeki tüm hocalarıma teşekkür ederim.

Öğrenim hayatım süresince her zaman yanımda olan ve desteklerini hiçbir zaman esirgemeyen annem Bilgi GÖKTÜRK'e, babam Ertuğrul GÖKTÜRK'e, kardeşim Erbil GÖKTÜRK'e, eşim Eren ŞAHİN'e, bu arada bana gülcükler atarak stresimi alan oğlum Ömer ŞAHİN'e ve arkadaşlarım Ceren SÜMER ve Tuğba KARAN'a sonsuz teşekkürlerimi sunarım.

Ersegül GÖKTÜRK ŞAHİN
Trabzon 2014

TEZ BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Bulanık Kümelemeye Dayalı Parametre Tahmini” başlıklı bu çalışmayı baştan sona kadar danışmanım Doç. Dr. Türkan ERBAY DALKILIÇ’ın sorumluluğunda tamamladığımı, verileri kendim topladığımı, analizleri ilgili laboratuarlarda yaptığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 24/01/2014

Ersegül GÖKTÜRK ŞAHİN

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ BEYANNAMESİ	IV
İÇİNDEKİLER	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ	X
SEMBOL VE KISALTMALAR DİZİNİ.....	XI
1. GENEL BİLGİLER	1
1.1. Giriş	1
1.2. Önceki Çalışmalar	2
1.3. Bulanık Mantık Kavramı	4
1.4. Bulanık Küme	5
1.4.1. Bulanık Kümelerin Gösterimi	5
1.4.2. Üyelik Fonksiyonu	6
1.4.2.1. Üçgen Üyelik Fonksiyonu	6
1.4.2.2. Yamuk Üyelik Fonksiyonu.....	7
1.4.2.3. Gaussian Üyelik Fonksiyonu.....	8
1.5. Bulanık Kümelerde Temel Kavramlar ve İşlemler	9
1.5.1. Destek Kümesi.....	9
1.5.2. α – Kesmesi	9
1.5.3. Yükseklik.....	10
1.5.4. Normallik.....	10
1.5.5. Cebirsel Toplam	10
1.5.6. Kesişim	11
1.5.7. Birleşim	11
1.5.8. Tümlleme.....	12
1.6. Kümeleme Analizi.....	12

1.6.1.	Hiyerarşik Kümeleme	14
1.6.2.	Hiyerarşik Olmayan Kümeleme Yöntemleri	15
1.6.2.1.	K-Ortalamalar	15
1.6.2.2.	Bulanık Kümeleme	16
1.6.2.2.1.	Bulanık C-Ortalamalar Algoritması	17
1.6.2.2.2.	Gustafson-Kessel Algoritması	18
1.6.2.2.3.	Gath-Geva Algoritması	20
1.7.	Regresyon Çözümlemesi	21
1.8.	Aykırı Değerlerin Belirlenmesi	24
2.	YAPILAN ÇALIŞMALAR	26
2.1.	Bulanık Üyeliklere Dayalı Parametre Tahmini	26
2.2.	Uygulama	28
4.	SONUÇLAR VE ÖNERİLER	44
5.	KAYNAKLAR	45
ÖZGEÇMİŞ		

Yüksek Lisans Tezi

ÖZET

BULANIK KÜMELEMeye DAYALI PARAMETRE TAHMİNİ

Ersegül GÖKTÜRK ŞAHİN

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü

İstatistik ve Bilgisayar Bilimleri Anabilim Dalı
Danışman: Doç. Dr. Türkan ERBAY DALKILIÇ
2014, 47 Sayfa

Regresyon çözümlemesinde verilerin farklı dağılımlara sahip olması durumu tahmin sürecinde klasik çözümlemelerin dışına çıkılmasını gerektirir. Böyle durumlarda bulanık çözümleme yöntemleri alternatif yöntemler olarak kendini göstermektedir. Bulanık regresyon çözümlemesinin önemli adımlarından biri veri setini meydana getiren kümelerin belirlenmesi ve bu kümelere yer alan verilerin tahmine katkılarının derecelerini belirleyecek üyelik derecelerinin elde edilmesidir. Bu çalışmada veri setlerinin üyelik derecelerinin belirlenmesi aşamasında Bulanık C-Ortalamalar (FCM) ve Gustafson-Kessel (G-K) kümeleme algoritmasından faydalanılmış, elde edilen üyelik derecelerine dayalı parametre tahmini için bir algoritma önerilmiş ve algoritmadan elde edilen tahminler literatürde yer alan mevcut yöntemlerden elde edilen tahminler ile karşılaştırılmıştır.

Anahtar Kelimeler: Bulanık C-Ortalamalar Algoritması (FCM), Gustafson-Kessel (G-K) Algoritması, Bulanık Regresyon Çözümlemesi.

Master Thesis

SUMMARY

PARAMETRES ESTIMATION BASED ON FUZZY CLUSTERING

Ersegül GÖKTÜRK ŞAHİN

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistic and Computer Science Graduate Program
Supervisor: Assoc. Prof. Türkan ERBAY DALKILIÇ
2014, 47 Pages

In regression analysis, the data has a different distribution requires to go beyond the classical resolutions during the estimation process. In such cases, fuzzy analysis methods are known as alternative methods. One of the important steps of the fuzzy regression analysis is determination of clusters which form the data set and obtaining membership degree which determine degree of contribution of the data in these clusters to estimation. In this study, at the stage of determination of membership degree of data sets, Fuzzy C-means and Gustafson-Kessel clustering algorithm were utilized. It is proposed an algorithm for fuzzy regression analysis depending membership degree to be obtained. The estimates obtained from this algorithm are compared with estimates obtained from existing methods.

Key Words: Fuzzy C-Means (FCM) Algorithm, Gustafson-Kessel (G-K) Algorithm, Fuzzy Regression Analysis.

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Üçgen üyelik fonksiyonu	7
Şekil 2. Yamuk üyelik fonksiyonunu	8
Şekil 3. Gaussian üyelik fonksiyonu	9
Şekil 4. Normallik	10
Şekil 5. \tilde{A} ve \tilde{B} bulanık kümelerinin kesişimi	11
Şekil 6. \tilde{A} ve \tilde{B} bulanık kümelerinin birleşimi	11
Şekil 7. \tilde{A} bulanık kümesinin tümleyeni	12
Şekil 8. Kümeleme tekniklerinin sınıflandırılması	14
Şekil 9. FCM algoritmasına göre kümelenmiş veriler	29
Şekil 10. G-K algoritmasına göre kümelenmiş veriler.....	29
Şekil 11. EKK yöntemi ile elde edilen model.....	30
Şekil 12. FCM yöntemi ile elde edilen modeller	30
Şekil 13. GK yöntemi ile elde edilen modeller	31
Şekil 14. FCM algoritmasına göre kümelenmiş veriler	33
Şekil 15. G-K algoritmasına göre kümelenmiş veriler.....	33
Şekil 16. EKK yöntemi ile elde edilen model.....	34
Şekil 17. FCM yöntemi ile elde edilen modeller	34
Şekil 18. GK yöntemi ile elde edilen modeller	35
Şekil 19. FCM algoritmasına göre kümelenmiş veriler.....	37
Şekil 20. G-K algoritmasına göre kümelenmiş veriler.....	37
Şekil 21. EKK yöntemi ile elde edilen model.....	38
Şekil 22. FCM yöntemi ile elde edilen modeller	38
Şekil 23. GK yöntemi ile elde edilen modeller	39
Şekil 24. FCM algoritmasına göre kümelenmiş veriler.....	41
Şekil 25. G-K algoritmasına göre kümelenmiş veriler.....	41
Şekil 26. EKK yöntemi ile elde edilen model	42
Şekil 27. FCM yöntemi ile elde edilen modeller	42
Şekil 28. GK yöntemi ile elde edilen modeller	43

TABLÖLAR DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Sürekli deęişkenler için sık kullanılan uzaklık ölçüleri	13
Tablo 2. Verilerin üç farklı yöntem ile tahminleri ve hataları.....	28
Tablo 3. Doğrusal regresyon modelleri ve hataları	31
Tablo 4. Verilerin üç farklı yöntem ile tahminleri ve hataları.....	32
Tablo 5. Doğrusal regresyon modelleri ve hataları	35
Tablo 6. Verilerin üç farklı yöntem ile tahminleri ve hataları.....	36
Tablo 7. Doğrusal regresyon modelleri ve hataları	39
Tablo 8. Verilerin üç farklı yöntem ile tahminleri ve hataları.....	40
Tablo 9. Doğrusal regresyon modelleri ve hataları	43

KISALTMALAR VE SEMBOLLER DİZİNİ

FCM	: Bulanık C-Ortalama (Fuzzy C-Means)
G-G	: Gath-Geva
G-K	: Gustafson-Kessel
HKT	: Hata Kareler Toplamı
$J(X,u,v)$: Amaç fonksiyonu
$J(x,v)$: Amaç fonksiyonu
m	: Bulanıklık indeksi
max	: Maksimum
min	: Minimum
n	: Toplam nesne sayısı.
\hat{Y}_{EKK}	: En Küçük Kareler yöntemi kullanılarak Y 'nin tahmin edicisi
\hat{Y}_{FCM}	: Bulanık C-Ortalamalar algoritması kullanılarak Y'nin tahmin edicisi
\hat{Y}_{GK}	: Gustafson-Kessel algoritması kullanılarak Y'nin tahmin edicisi
$\mu_{\tilde{A}}(x)$: \tilde{A} Bulanık kümesindeki x verisinin üyelik derecesi.
F_i	: i. Kümenin bulanık kovaryans matrisi
p_i	: Önsel olasılık
x_i	: i.kümenin verisi
\hat{Y}	: Y'nin tahmin edicisi
ϵ_{EKK}	: En Küçük Kareler yöntemi kullanılarak hesaplanan hata terimi
ϵ_{FCM}	: Bulanık C-Ortalamalar algoritması kullanılarak hesaplanan hata terimi
ϵ_{GK}	: Gustafson-Kessel algoritması kullanılarak hesaplanan hata terimi
\tilde{A}	: A Bulanık Kümesi
\tilde{B}	: B Bulanık Kümesi
U_t	: t. iterasyondaki üyelik matrisi.
V_t	: t. iterasyondaki küme merkezi
u_{ij}	: i. kümenin j. elemanın üyelik derecesi
v_i	: i. kümenin merkezi
<	: Küçük
=	: Eşit

$>$: Büyük
\in	: Eleman
Σ	: Toplam
\cap	: Kesişim
\cup	: Birleşim
\leq	: Küçük ve eşit
\geq	: Büyük ve eşit
c	: Küme sayısı
$d(x_i, x_j)$: x_i ve x_j arasındaki uzaklık
ε	: Hata terimi ve iterasyon bitirme kriteri
e_i	: Artık
e_{is}	: Standartlaştırılmış artık
p_{ii}	: Gözlem uzaklığı
r_i	: Student artık

1. GENEL BİLGİLER

1.1. Giriş

Bulanıklık kavramı son yıllardaki gelişimiyle hayatımızın bir parçası haline gelmiştir. Bulanıklık belirsizliği ifade etmekte olup, insan düşüncesi bulanık kavramlarla doludur. ‘A köyü şehre çok uzak’, ‘Sınavım çok iyi geçti’, ‘Ömer çok zayıf ’ bulanık kavramlar içermektedir. Ne kadar uzak, ne kadar iyi ve ne kadar zayıf ? Bu belirsizlikler klasik mantık ile ifade edilemezler. Belirsizliği ölçmede üyelik fonksiyonları devreye girmektedir. Üyelik fonksiyonları [0,1] aralığında değerler alabilir ve toplamları bir’e eşit olmalıdır. Üyelik fonksiyonları yardımıyla veriler kümelenebilirler. Bir eleman tanımlanan kümelerden her birine belirli derecelerde üyedir. Örneğin; A köyü, çok uzak kümesine 0.70, uzak kümesine 0.28 ve yakın kümesine de 0.02 üyelik dereceleri ile ait olabilir. Böylece eleman her kümeye belirli derecelere ait olmaktadır. Birbirine çok benzeyen birimler aynı kümede benzer üyelik derecesine göre yer alırlar.

Bir veri setinde yer alan elemanların kümelenebilmesi işlemi, klasik kümeleme yöntemleri ile yapılabildiği gibi son yıllarda bulanık kümeleme işlemleri ile de yapılmaktadır. Bulanık kümeleme algoritmalarından en yaygın olan Bulanık C-Ortalamlar Algoritması (FCM), elemanların küresel şekilde, Gustafson-Kessel Algoritması (G-K), elemanların elipsoidal şekilde dağıldıklarında iyi sonuç veren algoritmalarlardır.

Regresyon analizinde, veri kümesinde yer alan gözlemlerin tek bir sınıftan geldiği varsayılır. Bağımlı değişken ve bağımsız değişken arasındaki basit doğrusal ilişki $Y = F(X) + \epsilon$ şeklinde ifade edilir. Regresyon çözümlemesinde verilerin farklı parametrelere sahip olması durumunda tahmin sürecinde klasik çözümlemelerin dışına çıkılmasını gerektirir. Böyle durumlarda bulanık çözümleme yöntemleri alternatif yöntemler olarak kendini göstermektedir. Bulanık regresyon çözümlemesinin önemli adımlarından biri veri setini meydana getiren kümelerin belirlenmesi ve bu kümelerde yer alan verilerin tahmine katkılarının derecelerini belirleyecek üyelik derecelerinin elde edilmesidir.

Bu çalışmada veri setinde yer alan gözlemlerin iki kümede toplanması durumunda, gözlemlerin bu kümelere ait olma durumları belirlenirken bulanık kümeleme yöntemlerinden, Bulanık C-Ortalamlar (FCM) ve Gustafson-Kessel (G-K) kümeleme algoritmalarından faydalanılmış bulanık üyelik derecelerine dayalı bir algoritma

önerilmiştir. Önerilen algoritma MATLAB de yazılan program ile işletilmiştir. Çözüm sürecinde veri setinde aykırı gözlemlerin bulunması durumu da dikkate alınarak yöntemin aykırı gözlemlere karşı duyarlılığı da irdelenmiştir. Elde edilen sonuçlara ilişkin hata miktarları, literatürde yer alan mevcut yöntemlerden En Küçük Kareler yöntemi ile elde edilen sonuçların hata miktarları ile karşılaştırılarak yöntemin kullanılabilirliği üzerinde durulmuştur.

1.2. Önceki Çalışmalar

Lotfi A. Zadeh (1965) bulanık kümelerle ilgili çalışmasında, bulanık mantık ve küme teorisini, bulanık kümelerde niteliklerin dereceli üyelik fonksiyonlarıyla ifade edildiği açıklamıştır.

Gustafson ve Kessel (1979) bulanık kümelemeyi kovaryans matrisi ile çözümlenmiş, elipsoidal veriler için Gustafson-Kessel algoritmasının tanımını vermişlerdir.

Bezdek, Ehrlich ve Full (1984) Bulanık C-Ortalamlar algoritması için Öklid, Diyagonal ve Mahalonobis uzaklık ölçülerini kullanmışlardır.

Takagi ve Sugeno (1985) bulanık ifadeler ve ilişkilerin kullanıldığı sistemlerin bulanık modellenmesi ve kontrolleri için uygulamalar vermişlerdir.

Dombi (1990) üyelik fonksiyonları üzerine yaptığı çalışmada farklı üyelik fonksiyonlarını tanımlamış, üyelik fonksiyonlarının varsayımlarını, özelliklerini ve matematiksel formlarını açıklamıştır.

Xie ve Beni (1991) bulanık kümeleme için bir geçerlilik kriteri sundukları bu çalışmada bir bulanık kümeleme algoritması vermişler ve önerdikleri geçerlilik fonksiyonunu S olarak tanımlamışlardır. S fonksiyonunun hesaplanmasında kullanılan ayrılma ve yoğunlaşma kriterlerinin tanımını vermişlerdir.

Chi-Bin ve Lee (1998) bulanık regresyon analizi için bulanık uyarlamalı ağlardan yararlanmış ve hata terimini hesaplamışlardır.

Raghu Krishnapuram ve Jongwoo Kim (1999) bulanık uyarlamalı kümeleme algoritması ve Gustafson-Kessel bulanık kümeleme algoritmasını karşılaştırmışlardır.

Chi-Bin ve Lee (2001) bulanık uyarlamalı ağ ile switching regresyon analizi adlı çalışmalarında bulanık sugeno çıkarsama sistemi ve bulanık uyarlamalı ağlardan faydalanarak regresyon modelinin bilinmeyen parametrelerini tahmini üzerinde çalışmış ve iki örnekle çalışmalarını desteklemişlerdir.

Tuan D. Pham (2001) olasılıklı bulanık c-ortalamlar algoritmasını açıklamış ve sayısal bir örnekle çalışmasını desteklemiştir.

Türkan Erbay Dalkılıç (2005) bağımsız değişkenlerin normal ve üstel dağılımdan gelmesi durumunda switching regresyonda bulanık sinir ağıları yaklaşımı ile parametre tahmini yapmış ve hata terimini, en küçük kareler yönteminden elde ettiği hata terimi ile karşılaştırmıştır.

Umut Avcı (2006) Dbscan kümeleme yöntemi ve bulanık kümeleme yöntemi ile yoğunluk tabanlı kümeleme düşüncesini birleştiren fuzzy joint points kümeleme metotlarını karşılaştırmıştır.

Nevin Güler (2006) bulanık kümeleme algoritmalarından, bulanık c-ortalamlar ve bulanık c-regresyon kümeleme algoritmalarının bulanık modellemeye uygulamalarını ele almıştır. Mamdani, Takagi-Sugeno ve Sugeno-Yasukawa gibi çok iyi tanınan bulanık modellerin yapısı ve kurulmaları için gerekli aşamaları açıklamış, bulanık modellemeyi kullanarak Türkiye'nin elektrik tüketiminin tahminini yapmıştır.

Rıdvan Saraçoğlu (2007) bulanık kümeleme kullanarak benzer belge aranması çalışmasını yayımlamıştır.

Yunus Kocatürk (2007) yenileme süreçlerinde bulanık değişkenleri kullanarak oluşan bulanık yenileme sürecinde, beklenen değer, elementer bulanık yenileme süreçleri ve bulanık yenileme ödül teoremini ele almıştır.

Furkan Başer (2007) sigorta şirketinin ayırması gereken hasar karşılığı tutarı belirlenmesi için geleneksel en küçük kareler regresyonunu kullanan London Chain Ladder yöntemine alternatif olarak melez bulanık en küçük kareler regresyon çözümlemesi yöntemini önermiştir.

Solmaz Ene (2009) kalp hastalıklarının yaşam kalitesi üzerine etkili faktörleri belirlemek için Tanaka tarafından geliştirilmiş bulanık doğrusal regresyon modelini kullanarak bulanık parametreleri tahmin etmiştir.

Necati Alp Erilli (2009) Devlet Planlama Teşkilatının, illerin gelişmişlik göstergelerinde kullanılan sosyo-ekonomik, sağlık, sanayi, tarım, eğitim, istihdam, mali veriler yardımıyla Türkiye'deki 81 ili kümelere ayırmıştır.

Dilara Büyüköz (2010) bulanık c-ortalamlar algoritmasını kullanarak Avrupa Birliği ülkeleri ve Türkiye'nin ekonomik, nüfus, cinsiyet ayrımcılığı, ulaşım altyapısı, teknolojik altyapı ve teknolojik gelişim değişkenlerine göre kümeleme yapmıştır.

Murat Ay (2010) bulanık c-ortalamlar algoritmasını ve bulanık sistemleri kullanarak sulama suyu kalitesini kümelemiştir.

Ayşe Merve Gün (2011) yüz tanınması işlemini bulanık c-ortalamlar algoritması ile sınıflandırma yaparak gerçekleştirmiştir.

Erhan Ergün (2011) bulanık küme teorisi ve bulanık doğrusal programlamayı ele almış. Prefabrik konut fiyatlarının tahminine yönelik sayısal bir modelle çalışmasını desteklemiştir.

E.Eğrioğlu, C.H.Aladağ, U. Yolcu, V.R. Uslu ve N.A. Erilli (2011) Gustafson-Kessel algoritmasına dayanan bulanık zaman serileri adlı çalışmalarını yayımlamışlardır.

Ainura Turusbekova (2012) k-ortalamlar, bulanık eksiltici algoritma ve bulanık c-ortalamlar algoritmasını kullanarak mamografik kitle sınıflandırma yapmıştır.

1.3. Bulanık Mantık Kavramı

1956 yılında Amerika Birleşik Devletinde yapılan bir konferansta bulanıklık kavramından söz edilmiştir. 1965 yılında California Berkeley Üniversitesinden Prof. A. Lotfi Zadeh Bulanık Mantık ve Bulanık Kümeler makalesini yayınlamasıyla düşüncelerini ortaya koymuştur.

Aristotelesin 3 temel ilkesi

1. Özdeşlik ilkesi: Bir şey ne ise odur.
2. Çelişmezlik ilkesi: Bir şey hem kendi, hem de başka bir şey olamaz.
3. Üçüncünün olmazlığı ilkesi: Bir şey ya A'dır ya da A olmayandır. Üçüncü bir durum düşünülemez (Baykal ve Beyan,2004).

Bulanık mantık, Aristoteles ile başlayan klasik mantığa dayanan özdeşlik, çelişmezlik ve üçüncünün olmazlığı ilkelerine karşı geliştirilmiştir. Bulanıklık matematiksel olarak çok değerlilik anlamına gelir. Klasik mantıkta ikili değer mevcuttur. Bulanık mantık insan mantığında olduğu gibi sadece kesin değerleri değil ara değerleri de baz almaktadır. Çok sıcak, sıcak, ılık, soğuk ve çok soğuk gibi dilsel terimler bulanık değişkenlerdir.

Klasik mantık, 30°C'yi "sıcak" kümesinin sınırı olarak kabul ediyorsa, 29,9 °C'yi sıcak olarak kabul etme hakkını kaybeder. Oysaki dünyamızda 30°C'nin sıcak olduğu durumda, 29,9°C'nin soğuk olduğu iddia edilemez. Bulanık mantık bu kesin sınırları kaldırır.

Bulanık mantığın genel özellikleri Zadeh tarafından aşağıdaki gibi ifade edilmiştir.

1. Bulanık mantıkta, kesin değerler dayanan düşünme yerine yaklaşık düşünme kullanılır.
2. Bulanık mantıkta her şey $[0,1]$ aralığında belirli bir derece ile gösterilir.
3. Bulanık mantıkta bilgi, büyük, küçük, çok, az gibi dilsel ifadeler şeklindedir.
4. Her mantıksal işlem bulanık olarak ifade edilebilir.
5. Bulanık mantık matematiksel modeli zor elde edilen sistemler için daha uygundur (Elmas,2003).

1.4. Bulanık Küme

Bulanık küme, bir elemanın farklı üyelik dereceleriyle birden fazla kümeye ait olmasına imkân sağlar. Yani bir eleman birden fazla bulanık kümenin elemanı olabilir. Üyelik fonksiyonları bir elemanın bir kümeye ne derecede ait olduğunu belirleyen fonksiyonlardır. Üyelik derecesinin 0 olması durumu, elemanın kümeye ait olmadığını, 1 olması durumu ise kesin olarak ait olduğunu gösterirken, 1'e yakın değerler elemanın yüksek derecede kümeye ait olduğunu, 0'a yakın değerler de ise düşük derecede kümeye ait olduğunu gösterir. Buradan da anlaşılacağı gibi klasik kümede üyelik fonksiyonları $\{0,1\}$ değerlerini alırken; bulanık kümede üyelik fonksiyonları $[0,1]$ aralığında değerler alabilir.

1.4.1. Bulanık Kümelerin Gösterimi

Bulanık bir küme, bir nesne ve bu nesnenin ilgili kümeye üyelik derecesini gösteren sıralı çiftlerle ifade edilir. $\tilde{A} = (X, \mu_{\tilde{A}}(X))$ biçimindeki her bir çifte, bulanık teklik denir. Bulanık teklikler $\mu_{\tilde{A}}(X)/X$ şeklinde de gösterilebilirler.

Evrensel kümenin sonlu olduğu durumda \tilde{A} bulanık kümesinin gösterimi:

$$\tilde{A} = \sum \mu_{\tilde{A}}(X_i)/X_i = \mu_{\tilde{A}}(X_1)/X_1 + \mu_{\tilde{A}}(X_2)/X_2 + \dots + \mu_{\tilde{A}}(X_m)/X_m \quad (1)$$

ile verilir.

Evrensel kümenin sonsuz olduğu durumda ise \tilde{A} bulanık kümesinin gösterimi aşağıdaki gibidir:

$$\tilde{A} = \int \mu_{\tilde{A}}(X_i)/X_i \quad (2)$$

Yukarıda verilen ifadelerde Σ , \int , $/$ ve $+$ işaretleri sırasıyla cebirsel anlamda toplam, integral, bölme ve toplama işlemlerini göstermez. Σ ve \int işaretleri, bulanık tekliklerin kesikli ve sürekli evrenlerde bir araya getirilmesini ifade eder. $'/'$ simgesi bulanık teklikleri birbirinden ayırmaya yarayan bir ayıraçtır. $'+'$ işareti ise bulanık tekliklerin bir araya geldiğini göstermektedir (Ergün, 2011).

1.4.2. Üyelik Fonksiyonu

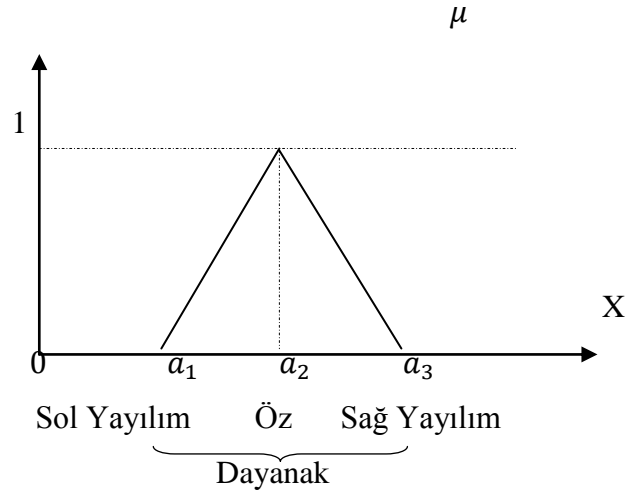
Üyelik fonksiyonları, bir elemanın bir kümeye ait olma derecesini belirleyen fonksiyonlardır. Literatürde çok sayıda üyelik fonksiyon mevcuttur. Bu çalışmada en yaygın olan üçgen üyelik fonksiyonu, yamuk üyelik fonksiyonu ve Gaussian üyelik fonksiyonu üzerinde durulacaktır.

1.4.2.1. Üçgen Üyelik Fonksiyonu

Üçgen üyelik fonksiyonu üç parametreden oluşur. Bunlardan a_1 üçgenin sol ayağını, a_3 sağ ayağını, a_2 tepe noktasının iz düşümünü göstermektedir ve

$$\mu_{\tilde{A}}(x, a_1, a_2, a_3) = \begin{cases} (x - a_1)/(a_2 - a_1), & a_1 \leq x \leq a_2 \\ (a_3 - x)/(a_3 - a_2), & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \text{ ve } x < a_1 \end{cases} \quad (3)$$

biçiminde ifade edilir.



Şekil 1. Üçgen üyelik fonksiyonu

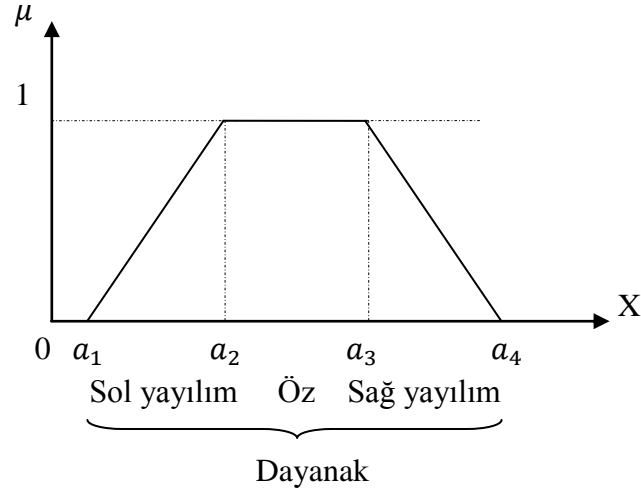
Şekil 1'deki X değişkeninin a_1 alt ve a_3 üst sınırları arasındaki her noktasına ayrı bir üyelik derecesi atanmış olur. $[a_1, a_2]$ aralığına sol yayılım, $[a_2, a_3]$ aralığına sağ yayılım denir. Eğer üçgen üyelik fonksiyonu simetrikse, $[a_1, a_2]$ ve $[a_2, a_3]$ aralıklarına yarıçap denir. $\mu(x) = 1$ değerini alan elemana, öz eleman denir. $\mu(x) > 0$ değerini alan elemanlar kümesinin dayanak (support) kümesini oluşturur.

1.4.2.2. Yamuk Üyelik Fonksiyonu

Bir yamuk üyelik fonksiyonu a_1, a_2, a_3, a_4 ile gösterilen dört parametreden oluşur ve

$$\mu_{\tilde{A}}(x, a_1, a_2, a_3, a_4) = \begin{cases} (x - a_1)/(a_2 - a_1), & a_1 \leq x \leq a_2 \\ 1, & a_2 \leq x \leq a_3 \\ (a_4 - x)/(a_4 - a_3), & a_3 \leq x \leq a_4 \\ 0, & x > a_4 \text{ ve ya } x < a_1 \end{cases} \quad (4)$$

biçiminde tanımlanır.



Şekil 2. Yamuk üyelik fonksiyonunu

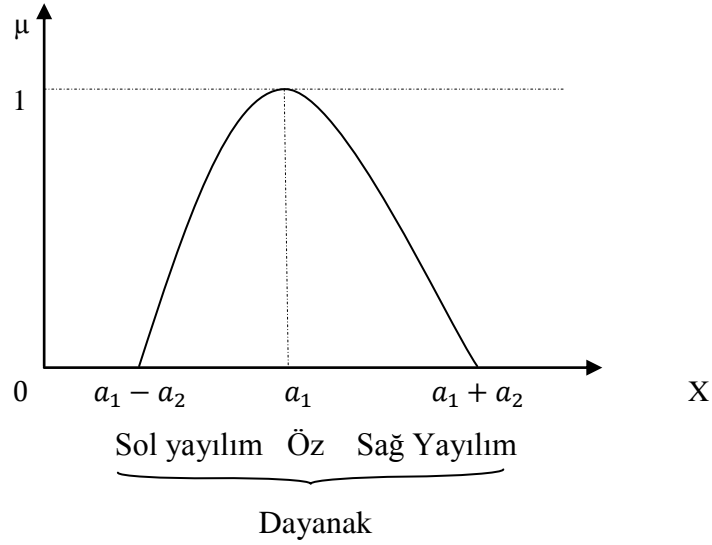
Şekil 2’de görüldüğü gibi, $[a_1, a_2]$ aralığına sol yayılım, $[a_3, a_4]$ aralığına sağ yayılım denir. $\mu(x) = 1$ değerini alan elemanlara öz eleman denir. $\mu(x) > 0$ değerini alan elemanlar kümenin dayanak (support) kümesini oluşturur.

1.4.2.3. Gaussian Üyelik Fonksiyonu

Gaussian üyelik fonksiyonu 2 parametreden oluşur. Bu parametreler a_1 ve a_2 ’ dir. “ a_1 ” merkezi temsil ederken “ a_2 ” ise merkezden sapmayı temsil etmektedir. Yani “ a_2 ” ne kadar büyük olursa üyelik fonksiyonu o kadar genişlemiş olur. Gaussian üyelik fonksiyonu

$$\mu_A(x) = \exp\left(-\frac{(x - a_1)^2}{2a_2^2}\right) \quad (5)$$

biçiminde ifade edilir.



Şekil 3. Gaussian üyelik fonksiyonu

Şekil 3'teki $[a_1 - a_2; a_1]$ aralığına sol yayılım, $[a_1 + a_2; a_1]$ aralığına sağ yayılım denir. $\mu(x) = 1$ değerini alan elemana öz eleman denir. $\mu(x) > 0$ değerini alan elemanlar kümenin dayanak (support) kümesini oluşturur.

1.5. Bulanık Kümelerde Temel Kavramlar ve İşlemler

1.5.1. Destek Kümesi

Destek (dayanak, support) kümesi, üyelik derecesi sıfırdan farklı olan elemanların bir araya getirildiği kümeye denir. Matematiksel gösterimi aşağıdaki gibidir;

$$\text{Support}(\tilde{A}) = \{x \in X, \mu_{\tilde{A}}(x) > 0\} \quad (6)$$

1.5.2. α - Kesmesi

Bulanık bir kümenin α - kesmesi, üyelik fonksiyon değeri α ' ya eşit ve daha büyük olan elemanların yer aldığı kümeye denir. α değeri, $\alpha \in (0,1]$ koşuluyla tanımlanan gerçel bir sayıdır. Matematiksel gösterimi aşağıdaki gibidir;

$$\tilde{A}_{\alpha} = \{x \in X, \mu_{\tilde{A}}(x) \geq \alpha\} \quad (7)$$

1.5.3. Yükseklik

Bulanık bir kümenin en büyük üyelik derecesi kümenin yüksekliğini belirler. Matematiksel gösterimi:

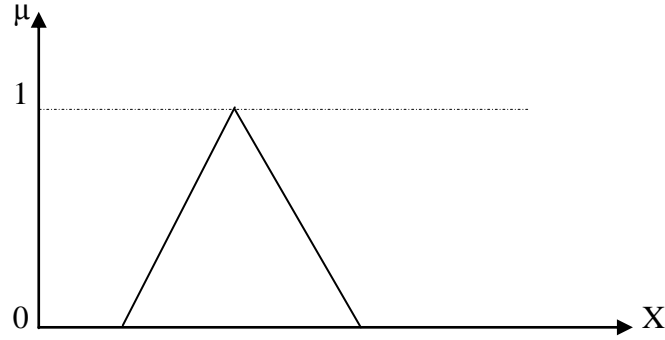
$$\text{Yükseklik } (\tilde{A}) = \{\mu_{\tilde{A}}(x), x \in X\} \quad (8)$$

1.5.4. Normallik

Bulanık bir kümenin en büyük üyelik derecesi 1 ise küme normaldir. Matematiksel gösterimi:

$$\text{Yükseklik } (\tilde{A}) = \sup\{\mu_{\tilde{A}}(x) = 1, x \in X\} \quad (9)$$

biçimindedir ve Şekil 4'te gösterilmiştir.



Şekil 4. Normallik

1.5.5. Cebirsel Toplam

\tilde{A} ve \tilde{B} iki bulanık küme olsun. $\tilde{A} + \tilde{B}$ 'nin üyelik fonksiyonu ile gösterimi

$$\mu_{\tilde{A} + \tilde{B}}(x) = \{\mu_{\tilde{A}}(x) + \mu_{\tilde{B}}(x), x \in X\} \quad (10)$$

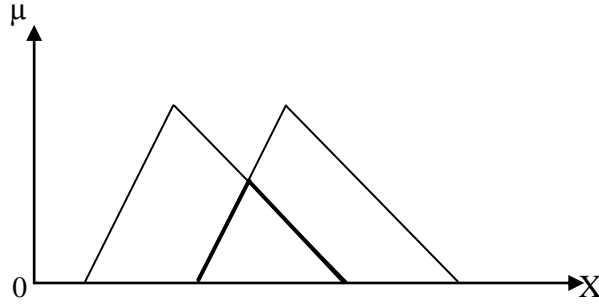
ile verilir.

1.5.6. Kesişim

\tilde{A} ve \tilde{B} iki bulanık küme olsun. $\tilde{A} \cap \tilde{B}$ 'nin üyelik fonksiyonu ile gösterimi

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x), x \in X\} \quad (11)$$

ile verilir.



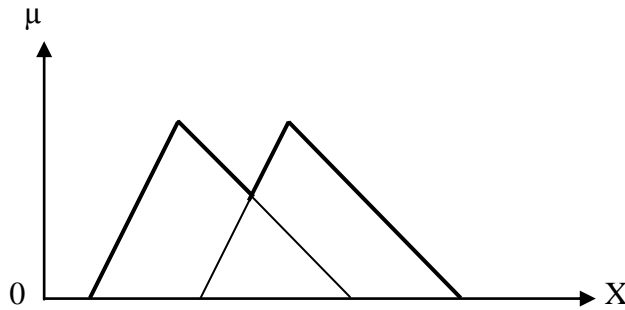
Şekil 5. \tilde{A} ve \tilde{B} bulanık kümelerinin kesişimi

1.5.7. Birleşim

\tilde{A} ve \tilde{B} iki bulanık küme olsun. $\tilde{A} \cup \tilde{B}$ 'nin üyelik fonksiyonu ile gösterimi

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = \max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x), x \in X\} \quad (12)$$

ile verilir.



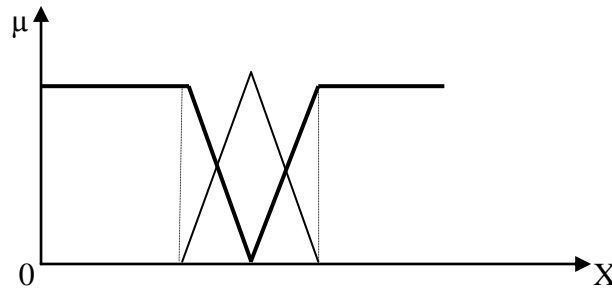
Şekil 6. \tilde{A} ve \tilde{B} bulanık kümelerinin birleşimi

1.5.8. Tümlleme

A bulanık küme olsun. A'nın üyelik fonksiyonu ile gösterimi

$$\mu_{\tilde{A}}(x) = \{1 - \mu_A(x), x \in X\} \quad (13)$$

ile verilir.



Şekil 7. \tilde{A} bulanık kümesinin tümlenyeni

1.6. Kümeleme Analizi

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlar yapmayı sağlayan bir yöntemdir.

Kümeleme analizi, elde bulunan veri yığınına belirlenen yöntemlerle analiz ederek daha önceden etiketleri belli olmayan gruplara ayrılması işlemidir. Bu işlem sonucunda elde edilen kümeler yüksek düzeyde küme içi homojenlik (aynı tür olan) ve kümeler arası heterojenlik (farklı tür olan) gösterirler (Atbaş,2008).

Kümeleme analizinin adımları aşağıdaki gibi verilebilir;

Adım 1: Değişkenlerin doğal gruplamaları hakkında kesin bilgilerin bulunmadığı popülasyonlardan alınan n sayıda birimin p sayıda değişkenine ilişkin gözlem değerleri elde edilir.

Adım 2: Birimlerin birbirleri ile olan benzerliklerini ya da farklılıklarını gösteren uygun bir benzerlik ölçüsü ile birimlerin birbirlerine uzaklıkları hesaplanır.

Adım 3: Uygun kümeleme yöntemi kullanılarak benzerlik/farklılık matrisine göre birimler uygun sayıda kümelere ayrılır (Güler,2006).

Kümeleme analizi, iki gözlemin benzerlikleri veya farklılıkları temel alınarak yapılır. Uzaklık ölçüleri ya da benzerlik ölçüleri veri matrisinde yer alan değişkenlerin ölçü birimlerine göre de farklılık göstermektedir (Güler,2006).

Kümeleme analizinde kullanılan benzerlik ölçütleri aşağıdaki gibi verilebilir;

1. Manhattan (City-Block) Uzaklığı
2. Öklit Uzaklığı
3. Ölçekli Öklit Uzaklığı
4. Mahalanobis Uzaklığı
5. Canberra Uzaklığı
6. Minkowski Uzaklığı

Kümeleme Analizinde en çok kullanılan uzaklık ölçütleri Öklit uzaklığı ve Mahalanobis uzaklığıdır. Uzaklık ölçütleri ve kullanılan uzaklık hesaplama formülleri Tablo 1’de yer almaktadır.

Tablo 1. Sürekli değişkenler için sık kullanılan uzaklık ölçüleri

Uzaklık Ölçüleri	Formül
Öklid uzaklığı	$D(x_i, x_j) = \sqrt{\sum_{i=1}^n (x_{ik} - x_{jk})^2}$
Ölçekli Öklit Uzaklığı	$D(x_i, x_j) = \sum_{i=1}^n (x_{ik} - x_{jk})^2$
Manhattan (City-Block) Uzaklığı	$D(x_i, x_j) = \sum_{i=1}^n x_{ik} - x_{jk} $
Minkowski Uzaklığı	$D(x_i, x_j) = \sqrt[p]{\left(\sum_{i=1}^n (x_{ik} - x_{jk})^p\right)}$
Mahalanobis Uzaklığı	$D(x_i, x_j) = (x_i - x_j)' F^{-1} (x_i - x_j)$
Canberra Uzaklığı	$D(x_i, x_j) = \sum_{i=1}^n \frac{ x_{ik} - x_{jk} }{ x_{ik} + x_{jk} }$

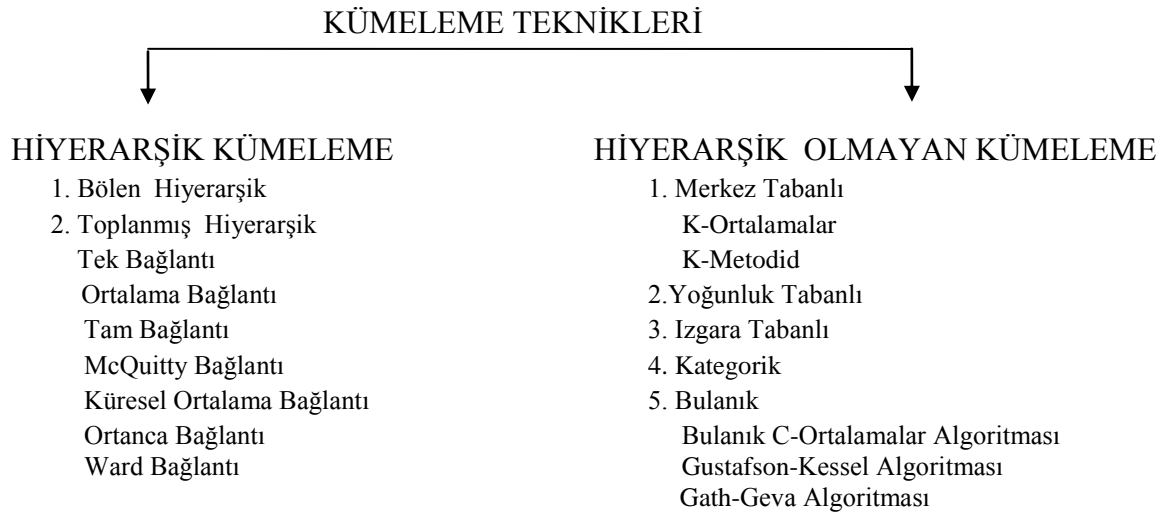
Öklid uzaklığı çoğunlukla iki ya da üç boyutlu uzaydaki nesnelerin birbirlerine olan yakınlıklarını ölçmek için kullanılır. Veri kümesi sıkı ya da izole edilmiş kümeler içerdiğinde genellikle iyi sonuçlar verir. Küresel küme şekillerini bulmaya eğilimlidir.

Minkowski uzaklık ölçüsü ise en büyük ölçekli nesnenin diğerlerine baskın gelmesi mantığına dayanır. Formülde de görüldüğü gibi, Minkowski ölçüsünün $p=2$ durumu Öklid uzaklığını, $p=1$ durumu ise Manhattan uzaklığını vermektedir. Öklid uzaklığı ölçüsü çok kümelenmeye neden olur ki, bu kümelerin sabit üyelik yüzeyleleri çok boyutlu küresel kümelerdir.

Mahalanobis ölçüsü ise uzaklık ölçüsünde kovaryansları (F) dikkate aldığı için, farklı ölçü birimlerine sahip veriler arasındaki uzaklıkları hesaplamak için uygundur. Mahalanobis uzaklık ölçüsünde çok boyutlu elips kümelerin eksenleri x koordinat eksenleriyle bir açı oluşturarak farklı şekillerde biçimlendirilebilir.

Ayrıca nesnelar arasında bir doğrusal ilişki olması da, uzaklık ölçülerini etkileyebilecektir. Bu durumlarda veri kümesinin normalleştirilmesi ya da Mahalanobis uzaklığının karesinin kullanılması ile daha doğru sonuçlar elde edilebilir (Büyükköz,2010).

Kümeleme analizi yaparken kullanılan teknikler şekil 8'de verildiği gibi sınıflandırılabilir;



Şekil 8. Kümeleme tekniklerinin sınıflandırılması

1.6.1. Hiyerarşik Kümeleme

Hiyerarşik kümeleme yöntemi, Bölen Kümeleme Yöntemi ve Toplanmış Hiyerarşik Kümeleme Yöntemi olarak iki ana başlıkta incelenir. Bölen Kümeleme Yöntemi başlangıçta tüm birimlerin bir küme oluşturduğunu kabul ederek birimleri hiyerarşik n birimi sırasıyla 1, 2, 3,..., n-1, n kümeyle ayırmayı amaçlayan bir yöntemdir. Toplanmış

hiyerarşik kümeleme yönteminin tersidir. Toplanmış yöntemle ilişkin sonuçlardan bölen hiyerarşik yöntemle ilişkin sonuçlar elde edilebilir (Yıldırım,2007).

Toplanmış Hiyerarşik Kümeleme yöntemi ise tek noktadan oluşan kümelerden başlar ve bu kümeleri tekrarlı olarak en uygun iki veya daha fazla kümede birleştirir. Başlangıçta tüm birimleri aynı kümede kabul eder ve n tane birimi sırasıyla n, n-1, n-2, ... , 3, 2, 1 kümeyle yerleştirmeyi amaçlar (Erilli,2009).

1.6.2. Hiyerarşik Olmayan Kümeleme Yöntemleri

Hiyerarşik olmayan kümeleme yöntemleri, merkez tabanlı, yoğunluk tabanlı, ızgara tabanlı, kategorik ve bulanık kümeleme başlıkları altında toplanabilirler. Merkez tabanlı kümeleme yöntemleri K-Ortalamlar yöntemi, bulanık kümeleme yöntemlerine temel teşkil etmesi bakımından önemlidir.

1.6.2.1. K-Ortalamlar

K-Ortalamlı kümeleme algoritması, hata kareler kriterini kullanan kesin kümeleme tekniğidir. Hata kareler toplamı (HKT)

$$HKT = \sum_{i=1}^N \sum_{r \in i} \|x_r^i - v_i\|^2 \quad (14)$$

ile verilir.

HKT nesnelere buldukları kümenin merkez noktalarına olan uzaklıklarının kareleri toplamını vermektedir. Nesnelere, buldukları küme merkezlerine olan konumları HKT değerinin yüksek ya da düşük çıkmasına sebep olur. Düşük HKT değeri nesnelere, küme merkezlerine yakın konumlandıkları, yüksek HKT değeri ise nesnelere küme merkezlerinin birbirlerine uzak oldukları anlamına gelmektedir. Amaç sıkı, birbirine benzer yani yakın nesnelere barındıran kümeler elde etmek olduğundan en düşük HKT değerine sahip kümeleme en iyi sonucu vermektedir.

K-Ortalamlar algoritması da kümeleme işlemi sonunda, mümkün olan en sıkı ve en ayrı k tane kümenin elde edilmesini hedefler. Rassal seçilen bir bölümeleme ile başlar ve nesnelere küme merkezleri arasındaki uzaklığa bağlı olarak, hata kriteri sağlanana kadar

nesnelerin kümelere devri gerçekleştirilir. Yani x_1, x_2, \dots, x_N her biri p değişkenli gözlem vektörleri, çok boyutlu bir X uzayında bir nokta olarak düşünüldüğünde ve aynı anda v_1, v_2, \dots, v_k her gözlem grubu için küme merkezi olarak seçildiğinde;

$$W_N = \frac{1}{N} \sum_{i=1}^N \min_{1 \leq j \leq k} \|x_i - v_j\|^2 \quad (15)$$

formülü kullanılarak nesnelere en yakın kümeye atanmaktadır. Küme sayısını gösteren k parametresi, kullanıcı tarafından algoritmanın çalıştırılmasından önce belirlenir.

K-ortalamlar algoritması aşağıdaki gibi tanımlanabilir;

Adım 1: K tane başlangıç küme merkezi olarak rastgele seçilir.

Adım 2: Küme merkezi olmayan birimler belirlenen uzaklık ölçülerine göre başlangıç küme merkezlerinin ait oldukları kümelere atanır.

Adım 3: Yeni küme merkezleri oluşturulan k adet başlangıç kümesindeki değişkenlerin ortalamaları alınarak tekrar oluşturulur.

Adım 4: Birimler, en yakın oldukları yeni küme merkezli kümelere atanır. Uzaklıklar yeniden hesaplanır.

Adım 5: Bir önceki küme merkezlerine olan uzaklıklar ile yeni oluşturulan küme merkezlerine olan uzaklıklar karşılaştırılır. Uzaklıklar makul görülebilir oranda azalmış ise 4. adıma dönlür. Eğer çok büyük bir değişiklik söz konusu değilse ise, iterasyon sona erdirilir. En son kullanılan merkezler optimal küme merkezleri olarak belirlenir.

1.6.2.2. Bulanık Kümeleme

Kümeleme analizinin hedefi, uygun benzerlik ölçüsüne göre homojen sınıfların sayısına bağlı olarak, örneklerin kümesini alt kümelere ayırmaktır. Sınıflardan birine ait olan örnekler benzerdir. Ayrı sınıfların örnekleri mümkün olduğu kadar farklıdır. Klasik kümeleme analizinde farklı sınıfların sınırları kesindir, yani bir örnek sadece bir sınıfa aittir. Ancak pratikte bazı durumlarda verilerin ait olabileceği sınıfın sınırları kesin olarak tanımlanamayabilir. Bu durumda, bir gözlem bir ya da daha fazla sınıfa farklı üyelik dereceleri ile ait olabilirler. Verilerin ait oldukları sınıfların sınırları kesin olarak tanımlanamadığında, veri seti kümelere ayrılmak istendiğinde bulanık kümeleme için

tanımlanmış algoritmalarından faydalanılabilir. Alt kesimlerde bulanık kümeleme algoritmalarına değinilecektir (Erbay, 2005).

1.6.2.2.1. Bulanık C-Ortalamlar Algoritması

Bulanık C-Ortalamlar (FCM) algoritması bulanık kümeleme yöntemlerinin en iyi bilinen ve en yaygın kullanılanıdır. 1973 yılında Dunn tarafından ortaya atılmış, 1981'de Bezdek tarafından geliştirilmiştir. K-Ortalamlar kümeleme algoritmasının bulanık halidir. Algoritma amaç fonksiyonunu minimum yapmaya çalışır. Diğer algoritmalar gibi uzaklık ölçütlerinden yararlanır. Uzaklık ölçülerinden hangisinin kullanılacağı küme yapısına ve kullanılan algoritmaya bağlıdır. FCM algoritması Öklid uzaklığını kullanmaktadır. Böylece, FCM algoritması sonuçlandığında, p boyutlu uzaydaki noktalar küresel şekiller oluşturur. Bu nedenle algoritma en güzel sonucu küresel olarak dağılmış veriler için verir. Elipsoidal veriler için de kullanılabilen algoritmanın hata payı daha sonra üzerinde durulacak olan algoritmalara oranla daha büyük olacağı için bu şekilde dağılmış olan veriler için tercih edilmemelidir.

Veri setindeki bir nesnenin üyelik dereceleri, veri noktası ile küme merkezi arasındaki uzaklık hesaplanarak bulunur. Nesne hangi küme merkezine yakınsa o kümeye ait olma üyeliği, diğer kümelere ait olma üyeliğinden daha büyük olacaktır. FCM algoritmasının uygulanabilmesi için ilk olarak küme merkezlerinin veya üyelik derecesinin bilinmesi gerekir. Bu parametrelerin önceden bilinmesi zor olduğu için deneme yanılma yöntemi ile en uygun değerler elde edilebilir.

FCM algoritmasının amacı X veri setini, belirlenmiş c sayıda bulanık kümeye en optimal biçimde yerleştirmektir. Algoritma bu amacına, en küçük kareler tekniğinin genellemesi olan amaç fonksiyonunu yinelemeli olarak minimize ederek ulaşmaya çalışır.

Amaç fonksiyonu

$$J(u, v) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2 \quad (16)$$

biçiminde ifade edilir. Burada,

n : gözlem sayısını,

c : küme sayısını,

m : bulanıklık indeksini,

u_{ij} : i.kümenin j. elemanına ait üyelik derecesini,

v_i : i.kümenin merkezini

temsil etmektedir. Ağırlıklandırılmış en küçük kareler fonksiyonu olarak ifade edilen amaç fonksiyonu $J(u, v)$ ağırlıklandırılmış hata kareler toplamıdır.

Eşitlik (16) ile verilen amaç fonksiyonunun en küçüklenmesi prensibine dayanan FCM algoritması aşağıdaki gibi tanımlanabilir;

Adım 1: Başlangıç değerleri; küme sayısı c , bulanıklık indeksi m , küme merkezleri v_i ($i = 1, \dots, c$) ve durdurma kriteri ε belirlenir.

Adım 2: Başlangıç küme merkezleri v_i kullanılarak üyelik dereceleri matrisi

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{2/m-1}} \quad (i = 1, \dots, c) \quad (17)$$

ile hesaplanır.

Adım 3: Her küme için küme merkezleri

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (i = 1, \dots, c) \quad (18)$$

ile güncellenir.

Adım 4: $\|V_t - V_{t-1}\| < \varepsilon$ ise iterasyona son verilir. Aksi takdirde Adım 2'ye geri dönülür.

1.6.2.2.2. Gustafson-Kessel Algoritması

Gustafson ve Kessel (1979), bir veri setindeki farklı geometrik oluşumların incelenebilmesi için, standart FCM kümeleme tekniğinde Öklid uzaklığı yerine, uzaklık ölçümünde ilgili veri setlerinin kovaryans matrislerini dikkate alan Mahalonobis uzaklık

ölçüsü kullanarak yeni bir algoritma geliştirmişlerdir. Bu nedenle Gustafson-Kessel (GK) algoritması, standart FCM algoritmasının genişletilmiş halidir (Büyükköz, 2010).

Amaç fonksiyonu

$$J(X, u, v) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m (x_j - v_i)^t A_i (x_j - v_i) \quad (19)$$

biçiminde ifade edilir ve Gustafson Kessel algoritması aşağıdaki gibi tanımlanabilir;

Adım 1: Başlangıç değerleri; küme sayısı c , bulanıklık indeksi m , üyelik matrisi U ve durdurma kriteri ε belirlenir.

Adım 2: Her küme için küme merkezleri

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (i = 1, \dots, c) \quad (20)$$

ile hesaplanır.

Adım 3: Her küme için bulanık kovaryans matrisleri

$$F_i = \frac{\sum_{j=1}^n u_{ij}^m (x_j - v_i)(x_j - v_i)^t}{\sum_{j=1}^n u_{ij}^m} \quad (i = 1, \dots, c) \quad (21)$$

ile hesaplanır.

Adım 4: Her bir birey için Mahalanobis uzaklığı

$$D(x_i, v_i) = (x_i - v_i)^t A_i (x_i - v_i) \quad (i = 1, \dots, c) \quad (22)$$

ile hesaplanır. Burada;

$$A_i = \sqrt{\det(S_i)} S_i^{-1} \quad (23)$$

$$S_i = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m (x_j - v_i)(x_j - v_i)^t \quad (24)$$

biçimindedir.

Adım 5: Yeni üyelik değerleri matrisi

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{2/m-1}} \quad (i = 1, \dots, c) \quad (25)$$

ile belirlenir.

Adım 6: Yeni üyelik değerleri ile eski üyelik değerlerini karşılaştır. $\|U_t - U_{t-1}\| < \varepsilon$ ise iterasyona son verilir. Aksi takdirde Adım 2'ye geri dönlür.

1.6.2.2.3. Gath-Geva Algoritması

Gath ve Geva (G-G) Algoritması, kümelerin yoğunluklarını ve boyutlarını dikkate alan G-K algoritmasının geliştirilmiş bir versiyonudur. Yoğunluk odaklı çalıştığı için aykırı özellikler taşıyan veriler içeren yapılar için daha uygun bir algoritmadır. Gerçekte bu yaklaşım, amaç fonksiyonunu optimize etmeye dayanmaz. G-G algoritmasının temel prensibi, veri noktalarının p-boyutlu normal dağıldığını varsaymaktır. G-K algoritmasından farklı olarak, uzaklık ölçüsü üstel terim içermektedir ve kümeler hacimle sınırlandırılmamıştır. Bununla birlikte bu algoritma, üstel uzaklık ölçü biriminin kullanılması sebebiyle başlangıç koşullarına çok duyarlıdır. Küme merkezlerinin hesaplanması da G-K algoritması ile aynıdır (Büyükköz,2010).

Gath-Geva algoritması aşağıdaki gibi tanımlanabilir;

Adım 1: Başlangıç değerleri küme sayısı c , bulanıklık indeksi m , üyelik matrisi U ve durdurma kriteri ε belirlenir.

Adım 2: Her küme için küme merkezleri

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (i = 1, \dots, c) \quad (26)$$

ile hesaplanır.

Adım 3: Her küme için bulanık kovaryans matrisleri

$$F_i = \frac{\sum_{j=1}^n u_{ij}^m (x_j - v_i)(x_j - v_i)^t}{\sum_{j=1}^n u_{ij}^m} \quad (i = 1, \dots, c) \quad (27)$$

ile hesaplanır.

Adım 4: Önsel olasılık değeri

$$p_i = \frac{\sum_{j=1}^n u_{ij}^m}{\sum_{j=1}^n \sum_{i=1}^c u_{ij}^m} \quad (28)$$

ile hesaplanır.

Adım 5: Uzaklıklar

$$d(x, v) = \frac{1}{p} \sqrt{\det(A_i)} \exp\left(\frac{1}{2} (x_k - v_i)^t A_i^{-1} (x_k - v_i)\right) \quad (29)$$

ile hesaplanır.

Adım 6: Üyelik Matrisi

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|}\right)^{2/m-1}} \quad (i = 1, \dots, c) \quad (30)$$

ile hesaplanır.

Adım 7: $\|U_t - U_{t-1}\| < \varepsilon$ ise iterasyona son verilir. Aksi takdirde Adım 2'ye geri dönülür.

1.7. Regresyon Çözümlemesi

Regresyon çözümlemesinin genel amacı; bağımlı (açıklanan) değişken ile bağımsız (açıklayan) değişkenler arasındaki ilişkiyi matematiksel olarak modelleyerek, bağımsız değişkenin alacağı değerleri tahmin etmek şeklinde özetlenebilir.

Doğrusal regresyon modelinin parametrelerini tahmin etmek için en sık kullanılan yöntem En Küçük Kareler yöntemidir. En Küçük Kareler yönteminin varsayımları aşağıdaki gibi verilebilir.

1. Hata terimi rassaldır.
2. Hata teriminin ortalaması sıfırdır. $E(\varepsilon_i) = 0$
3. Hata terimleri sabit varyansa sahiptir. $\text{Var}(\varepsilon_i) = \sigma^2$

4. Hata terimleri sıfır ortalamalı, sabit varyanslı Normal Dağılıma sahiptir.

$$\varepsilon_i \sim N(0, \sigma^2)$$

5. Hata terimleri arasında ilişki yoktur.

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = E\left[[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)]\right] = 0$$

6. Bağımsız değişken ve hata terimi arasında ilişki yoktur.

$$\begin{aligned} \text{Cov}(\varepsilon_i, X_i) &= E[[\varepsilon_i - E(\varepsilon_i)][X_i - E(X_i)]] \\ &= E[\varepsilon_i(X_i - E(X_i))] \\ &= E(\varepsilon_i X_i) - E(\varepsilon_i) \cdot E(X_i) \\ &= E(\varepsilon_i, X_i) = 0. \end{aligned}$$

7. Bağımsız değişkenler arasında ilişki yoktur.

Bir bağımlı, bir bağımsız değişken olması durumunda basit doğrusal regresyon modeli;

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (31)$$

ile verilir.

Doğrusal EKK yönteminde esas kriter, regresyon hata terimleri karelerinin toplamını en küçük yapan $\hat{\beta}_0$ ile $\hat{\beta}_1$ değerlerini bulmaktır. O halde;

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad (32)$$

ifadesini en küçük yapacak $\hat{\beta}_0$ ve $\hat{\beta}_1$ değerini bulmak gerekecektir. Bu amaçla ifadenin önce $\hat{\beta}_0$ 'a, daha sonra $\hat{\beta}_1$ 'e göre birinci kısmi türevleri,

$$\frac{\partial S}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \quad (33)$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i \quad (34)$$

biçimindedir. Bu iki eşitlik sıfıra eşitlenip, gerekli işlemler yapıldıktan sonra normal denklemler olarak adlandırılan

$$\sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i \quad (35)$$

$$\sum_{i=1}^n X_i Y_i = \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 \quad (36)$$

biçimindeki denklemlere ulaşılır.

Bu iki denklemin eş anlamlı çözümünden $\hat{\beta}_1$ ve $\hat{\beta}_0$ tahmincileri

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (37)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (38)$$

biçiminde elde edilir.

Bağımsız değişkenin birden çok olması durumunda ise çoklu doğrusal regresyon modeli;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (39)$$

biçiminde tanımlanır.

Çoklu doğrusal regresyon modelinin matris ve vektör ile gösterimi;

$$Y = X\hat{\beta} + \varepsilon \quad (40)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

biçimindedir. Burada

Y: nx1 boyutlu bağımlı (açıklanan) değişken,

X: nxp boyutlu bağımsız (açıklayıcı) değişken,

β : px1 boyutlu regresyon katsayısı,

ε : nx1 boyutlu rassal hata terimidir.

Regresyon çözümlemesinde temel amacımız parametre tahminidir. Regresyon katsayıları β 'nin, tahmin edicilerini $\hat{\beta}$ şeklinde ifade edilir ve

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (41)$$

ile hesaplanır.

Bağımlı değişken Y 'nin tahmin edicisini \hat{Y} ile gösterilir.

$$\hat{Y} = X\hat{\beta} \quad (42)$$

ile belirlenir.

\hat{Y} değerleri belirlendikten sonra, ε hata terimleri

$$\varepsilon = (Y - \hat{Y}) \quad (43)$$

ile elde edilir.

En küçük kareler tahmin edicileri, gözlenen değerler ile tahmin değerleri arasındaki farkların kareleri toplamını minimize eder.

1.8. Aykırı Değerlerin Belirlenmesi

Bu çalışmada önerilen parametre tahmini yönteminin amaçlarından biri de veri setinde yer alan aykırı gözlemlerden mümkün olduğu kadar az etkilenen parametreler elde ederek hata miktarı düşük tahmin değerlerine ulaşmaktır. Bu anlamda veri setinde var olan aykırı gözlemlerin belirlenmesi önemlidir.

Regresyon çözümlemesindeki model bozukluklarının araştırmak ve aykırı değeri belirlemek için basit ve etkin yöntem artıkların incelenmesidir. Çünkü artıklar, hataların gerçekleşen ya da gözlenen değerleri olarak düşünülebilir. Bu nedenle hatalar için söz konusu olan varsayımlara ilişkin bozulmalar artıklar yardımı ile incelenir. Bu incelemede değişik artık türlerinden yararlanılır, ham artıklar, standartlaştırılmış artıklar, student türü artıklar bunlardan bazılarıdır (Alpar, 2003).

Klasik regresyon analizinden bilindiği gibi ham artık;

$$e_i = y_i - \hat{y}_i \quad i = 1, \dots, n \quad (44)$$

olarak tanımlanır. Bu tür artıklar ölçeklenmemiş artıklar olarak adlandırılır. Ham artıkların toplamı dolayısıyla ortalaması sıfır olmakla birlikte varyansı (S^2) örneklemeden örnekleme değişir. Varyansın değişiklik göstermesi sorunu artıkların standartlaştırılması ile giderilir. Bu durumda ham artıklar standart sapmaya (S) bölünerek standartlaştırılmış artıklar,

$$e_{is} = \frac{e_i}{S} \quad i = 1, \dots, n \quad (45)$$

biçiminde elde edilir. e_{is} artıklarının %95'inin çoğunlukla $[-2,+2]$ sınırları arasında değiştiği kabul edilir. Bu sınırların dışına çıkan gözlemlerin aykırı değer kuşkusu ile incelenmesi önerilir.

e_{is} artıkları her zamansız ortalama ve bir standart sapma ile normal dağılım göstermeyebilir. Bu durumda artıklar kendi standart sapması ile standartlaştırıldığında elde edilen artıklara Student türü artıklar denir ve

$$r_i = \frac{e_i}{\sqrt{S^2(1-p_{ii})}} \quad i = 1, \dots, n \quad (46)$$

biçiminde ifade edilir. Burada,

p_{ii} : gözlem uzaklığı değeridir, i . gözlem noktasının değişkenlerin oluşturduğu elipsoit merkezine uzaklığının bir ölçüsü olup 0 ile 1 arasında değişir ve

$$p_{ii} = X(X'X)^{-1}X' \quad (47)$$

ile ifade edildiğinde p_{ii} 'nin köşegen elemanları i . gözlem noktasına ilişkin gözlem uzaklığını verir. Bu uzaklıklarında standartlaştırılmış artıklarda olduğu gibi $[-2,+2]$ aralığında kalması beklenir. Aralık dışında kalan gözlemler artık gözlem olarak adlandırılır (Apaydın, 1994; Alpar, 2003).

2. YAPILAN ÇALIŞMALAR

Bu çalışmada bulanık kümeleme algoritmalarından elde edilen üyelik dereceleri kullanılarak bağımsız değişkenlerin modele katkıları ağırlıklandırılmış ve aykırı değerlerden mümkün olduğunca az etkilenen tahminler elde etmeyi amaçlayan bir algoritma önerilmiştir. Önerilen algoritmalarından elde edilen sonuçlar EKK yönteminden elde edilen sonuçlar ile karşılaştırılarak etkinliği irdelenmiştir.

2.1. Bulanık Üyeliklere Dayalı Parametre Tahmini

Doğrusal regresyon modelinin parametrelerinin belirlenmesi süreci verilerin dağılımının incelenmesi ile başlar. Veriler sadece bir kümede yığılma göstermediklerinde parametre tahmini yapılırken klasik yöntemlerin dışına çıkılması gerekir. Sürecin ilk adımı, veri setinde yer alan gözlemlerin kaç kümede toplanabileceklerini belirlemektir. Bu çalışmada basit doğrusal regresyon modeli üzerinde yoğunlaşıldığından küme sayısı saçılım grafikleri ile sezgisel olarak belirlenmiştir. Temel problem gözlemlerin hangi kümeye ait olduklarının ve ait olma üyelik derecelerinin belirlenmesidir. Bu anlamda kümeleme yapabilmek için Bulanık kümeleme yöntemlerine başvurulmuştur. Bulanık kümeleme yöntemlerinden en yaygın olanı Bulanık C-Ortalamlar algoritması, veriler küresel olarak dağıldığı zaman iyi sonuç veren bir algoritmadır. Bu algoritma veriler ile küme merkezi arasındaki uzaklığı ölçmek için Öklid uzaklık ölçüsünü kullanır. Gustafson-Kessel algoritması, FCM algoritmasının genişletilmiş olarak düşünülür ve elipsoidal olarak dağılan veriler için daha iyi sonuç verir. Uzaklık ölçüsü olarak mahalanobis uzaklık ölçüsünü kullanır. Doğrusal regresyon modelinin bilinmeyen parametrelerinin tahmininde bu iki kümeleme yöntemi esas alındığında bu yöntemlerden elde edilen bulanık üyeliklere dayalı parametre tahmini için algoritma aşağıdaki gibi tanımlanabilir;

Adım 1: G-K ve FCM kümeleme algoritmaları için küme sayısı c , bulanıklık indeksi m , durdurma kriteri ϵ belirlenir.

Adım 2: G-K kümeleme algoritması için optimal üyelik matrisi U eşitlik (25) kullanılarak belirlenir.

Adım 3: FCM kümeleme algoritması için optimal üyelik matrisi U eşitlik (17) kullanılarak belirlenir.

Adım 4: FCM ve G-K algoritmalarından elde edilen optimal üyelik matrisleri kullanılarak bağımsız değişken verileri her bir üyelik matrisi ile ayrı ayrı;

$$X_{W_{i(FCM)}} = u_{ij(FCM)}x_j \quad (i = 1, \dots, c)(j = 1, \dots, n) \quad (48)$$

$$X_{W_{i(GK)}} = u_{ij(GK)}x_j \quad (i = 1, \dots, c)(j = 1, \dots, n) \quad (49)$$

biçiminde ağırlıklandırılır.

Adım 5: Ağırlıklandırılmış veriler kullanılarak modelin parametreleri,

$$\hat{\beta}_{i(FCM)} = \left((x_i u_{ij(FCM)})' x_i \right)^{-1} \left((x_i u_{ij(FCM)})' Y_i \right) \quad (i = 1, \dots, c)(j = 1, \dots, n) \quad (50)$$

$$\hat{\beta}_{i(GK)} = \left((x_i u_{ij(GK)})' x_i \right)^{-1} \left((x_i u_{ij(GK)})' Y_i \right) \quad (i = 1, \dots, c)(j = 1, \dots, n) \quad (51)$$

biçiminde belirlenir.

Adım 6: \hat{Y}_i ile ifade edilen tahmin değerleri

$$\hat{Y}_{i(FCM)} = x_{i(FCM)} \hat{\beta}_{i(FCM)} \quad (i = 1, \dots, c) \quad (52)$$

$$\hat{Y}_{i(GK)} = x_{i(GK)} \hat{\beta}_{i(GK)} \quad (i = 1, \dots, c) \quad (53)$$

ile hesaplanır.

Adım 7: Kurulan modellere ilişkin tahminler ve gerçek değerleri arasındaki fark olarak tanımlanan hata miktarları, $e_i = (\hat{Y}_i - Y_i)$ ile belirlenir ve hata kareler toplamının gözlem sayısına oranı olarak belirlenen hata ölçütü

$$\varepsilon = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (54)$$

ile her model için hesaplanır. Küçük hataya sahip model sonuç model olarak kullanılır.

2.2. Uygulama

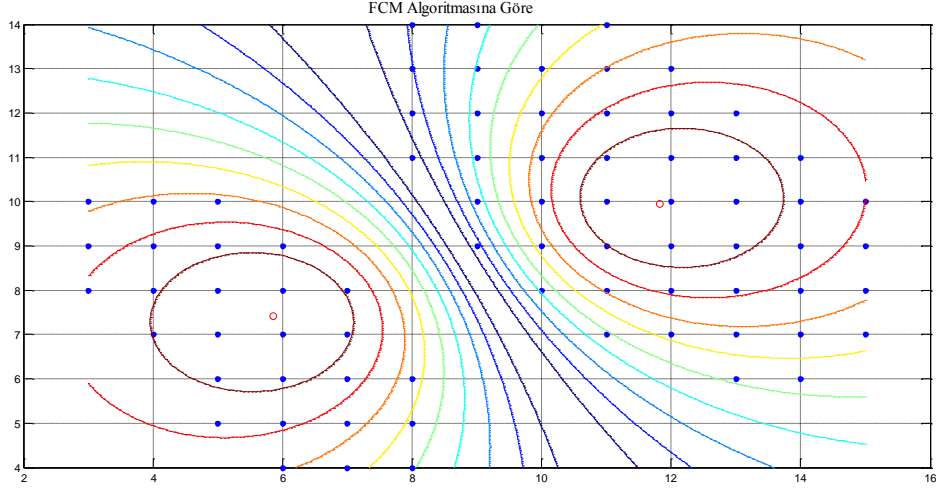
Bu bölümde verilerin birden fazla kümeye ayrılmaları durumunda bölüm (2.1) de önerilen algoritmanın kullanılabilirliğini göstermek amacı ile çeşitli uygulamalara yer verildi.

Örnek 1: Bu örnekte yer alan veri seti 75 gözlemden oluşmakta ve verilerin bir kısmı Tablo 2’de yer almaktadır. Veri setinde bir bağımlı ve bir bağımsız değişken bulunmaktadır. Bağımlı değişken Y, bağımsız değişken X ile ifade edilmiştir. Tablo 2’de ayrıca EKK yöntemi, FCM ve G-K algoritmalarından elde edilen tahminler ve bu tahminlere ilişkin hata miktarları yer almaktadır.

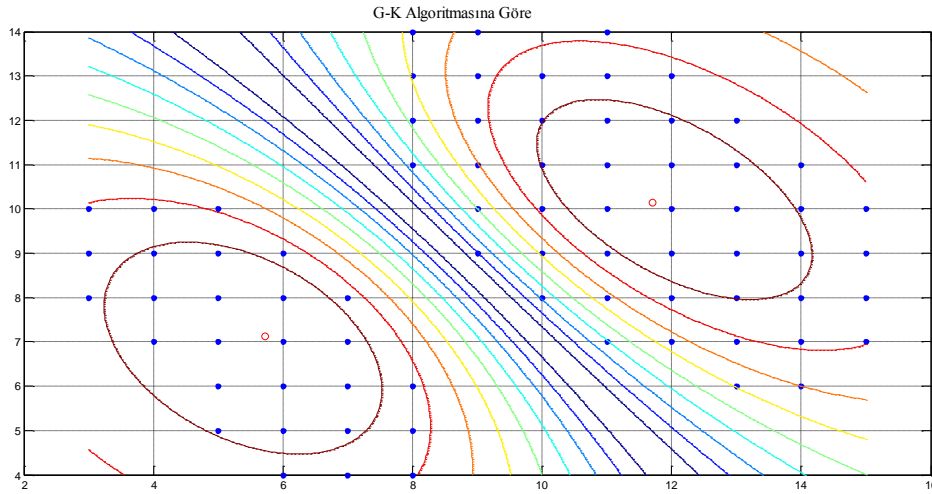
Tablo 2. Verilerin üç farklı yöntem ile elde edilen tahminleri ve hataları

X	Y	\hat{Y}_{EKK}	e_{EKK_i}	\hat{Y}_{FCM}	e_{FCM_i}	\hat{Y}_{GK}	e_{GK_i}
3	8	7.9284	0.0716	9.7854	-1.7854	8.4615	-0.4615
3	9	7.9284	1.0716	9.7854	-0.7854	8.4615	0.5385
3	10	7.9284	2.0716	9.7854	0.2146	8.4615	1.5385
4	7	8.0945	-1.0945	8.7100	-1.7100	7.9041	-0.9041
4	8	8.0945	-0.0945	8.7100	-0.7100	7.9041	0.0959
4	9	8.0945	0.9055	8.7100	0.2900	7.9041	1.0959
4	10	8.0945	1.9055	8.7100	1.2900	7.9041	2.0959
5	5	8.2605	-3.2605	7.6345	-2.6345	7.3467	-2.3467
5	6	8.2605	-2.2605	7.6345	-1.6345	7.3467	-1.3467
5	7	8.2605	-1.2605	7.6345	-0.6345	7.3467	-0.3467
5	8	8.2605	-0.2605	7.6345	0.3655	7.3467	0.6533
5	9	8.2605	0.7395	7.6345	1.3655	7.3467	1.6533
10	10	9.0907	0.9093	10.7761	-0.7761	10.6647	-0.6647
12	9	9.4228	-0.4228	9.8651	-0.8651	9.9002	-0.9002
12	10	9.4228	0.5772	9.8651	0.1349	9.9002	0.0998
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	9	9.9209	-0.9209	8.4986	0.5014	8.7534	0.2466
15	10	9.9209	0.0791	8.4986	1.5014	8.7534	1.2466
13	9	9.5888	-0.5888	9.4096	-0.4096	9.5179	-0.5179
12	11	9.4228	1.5772	9.8651	1.1349	9.9002	1.0998
6	6	8.4266	-2.4266	6.5591	-0.5591	6.7893	-0.7893
7	5	8.5926	-3.5926	5.4836	-0.4836	6.2319	-1.2319
6	7	8.4266	-1.4266	6.5591	0.4409	6.7893	0.2107
7	6	8.5926	-2.5926	5.4836	0.5164	6.2319	-0.2319
12	10	9.4228	0.5772	9.8651	0.1349	9.9002	0.0998
12	9	9.4228	-0.4228	9.8651	-0.8651	9.9002	-0.9002
11	10	9.2568	0.7432	10.3206	-0.3206	10.2824	-0.2824
10	11	9.0907	1.9093	10.7761	0.2239	10.6647	0.3353
10	12	9.0907	2.9093	10.7761	1.2239	10.6647	1.3353
10	9	9.0907	-0.0907	10.7761	-1.7761	10.6647	-1.6647
HATA		$\epsilon_{EKK} = 6.1691$		$\epsilon_{FCM} = 5.6438$		$\epsilon_{GK} = 5.7418$	

Şekil 9 ve 10'da bulanık kümeleme algoritmalarından FCM ve G-K algoritmalarına göre kümelenmiş veriler yer almaktadır.



Şekil 9. FCM algoritmasına göre kümelenmiş veriler

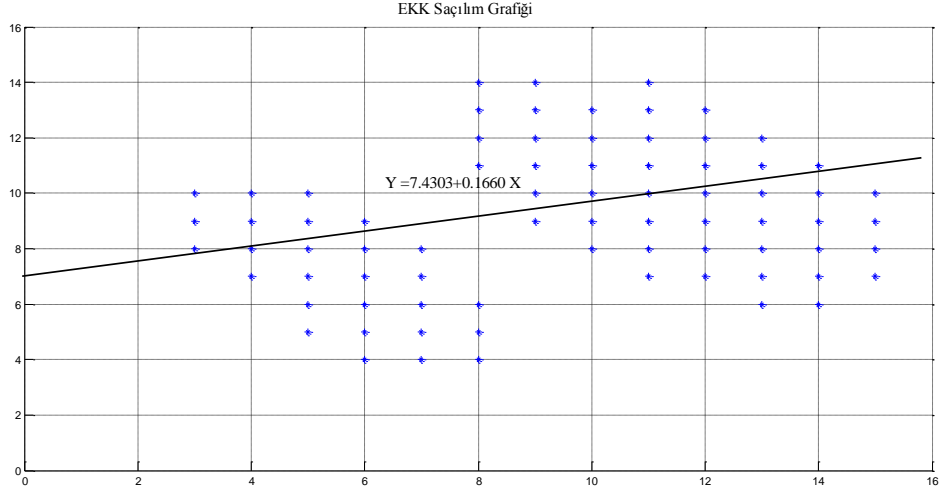


Şekil 10. G-K algoritmasına göre kümelenmiş veriler

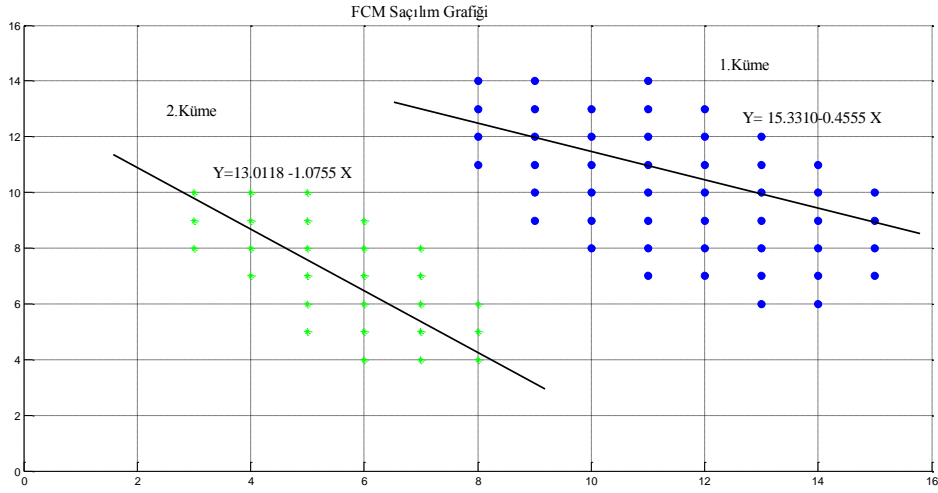
Şekil 11'de 75 verinin saçılım grafiğini ve En Küçük Kareler yöntemi ile elde edilen modeli yer almaktadır.

Şekil 12'de 75 veri Bulanık C-Ortalamlar yöntemi ile iki kümeye ayrılmış ve üyelik dereceleri kullanılması ile elde edilen modeller ayrı ayrı gösterilmiştir.

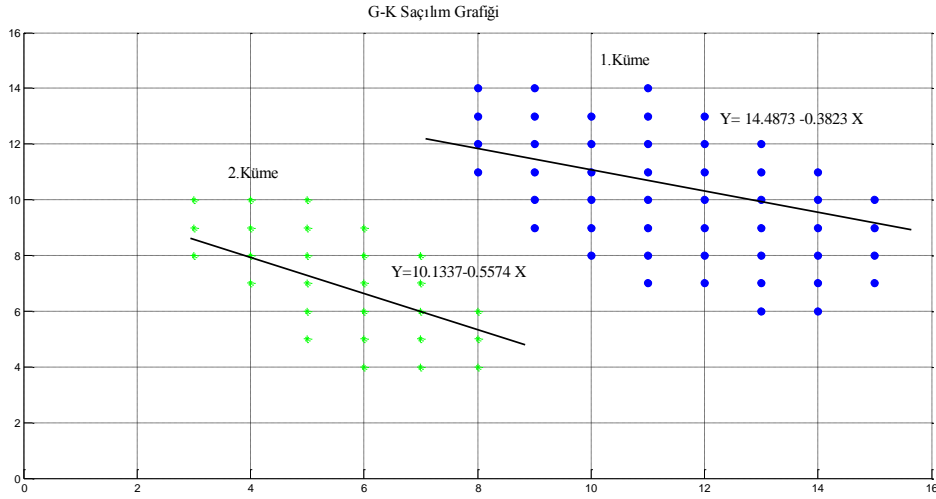
Şekil 13'te 75 veri Gustafson-Kessel yöntemi ile iki kümeye ayrılmış G-K algoritmasından elde edilen üyelik derecelerinin kullanılması ile oluşturulan modellere yer verilmiştir.



Şekil 11. EKK yöntemi ile elde edilen model



Şekil 12. FCM algoritması ile elde edilen modeller



Şekil 13. G-K algoritması ile elde edilen modeller

Tablo 3’te kullanılan üç yöntem ile elde edilen parametreler ile oluşturulan modeller yer almaktadır.

Tablo 3. Doğrusal regresyon modelleri ve hataları

En Küçük Kareler Yöntemi	Bulanık C-Ortalamalar	Gustafson-Kessel
$\hat{Y} = -7.4303 + 0.1660 X$	$\hat{Y}_1 = 15.3310 - 0.3823X_{1j}$ $\hat{Y}_2 = 13.0118 - 1.0755X_{2j}$	$\hat{Y}_1 = 14.4873 - 0.3823X_{1j}$ $\hat{Y}_2 = 10.1337 - 0.5574X_{2j}$
$\varepsilon_{EKK} = 6.1691$	$\varepsilon_{FCM} = 5.6438$	$\varepsilon_{GK} = 5.7418$

Bulanık ağırlıklara dayalı parametre tahmini yapılırken, FCM ile kümelenecek ve buradan elde edilen üyelik dereceleri kullanılarak ağırlıklandırılmış verilerden tahmin edilen parametreler ile kurulan modele ilişkin hata; eşitlik (54) ile verilen hata kriterlerinden 5.6438 olarak hesaplanmıştır. G-K ile kümelenecek ve buradan elde edilen üyelik dereceleri kullanılarak ağırlıklandırılmış verilerden tahmin edilen parametreler ile kurulan modele ilişkin hata; eşitlik (54) ile verilen hata kriterlerinden 5.7418 olarak hesaplanmıştır. EKK yöntemi ile elde edilen parametrelere ilişkin modelin hatası 6.1691 olarak belirlenmiştir.

Sonuç olarak FCM’ e dayalı modelden elde edilen tahminler en küçük hata miktarına sahiptir.

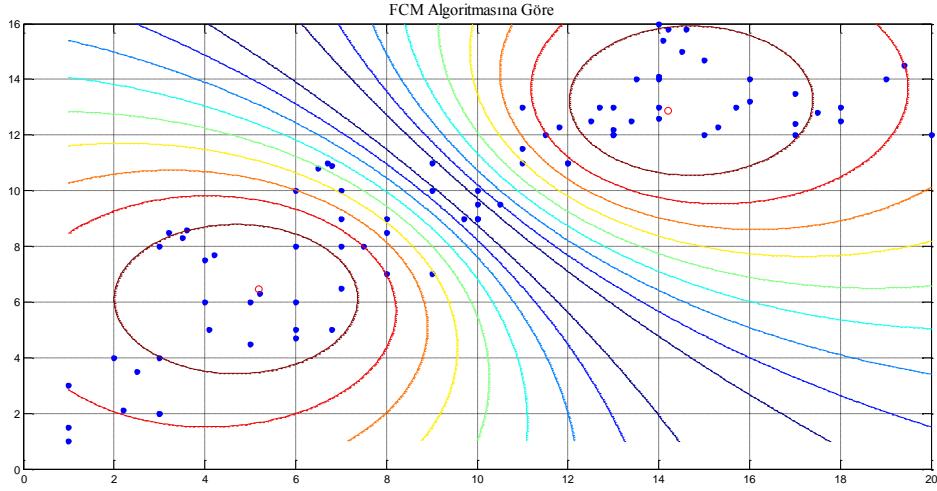
Örnek 2: Bu örnekteki veri seti 85 gözlemden oluşmakta ve verilerin bir kısmı Tablo 4'te yer almaktadır. Veri seti bir bağımlı ve bir bağımsız değişkene ilişkin gözlemlerden oluşmaktadır. Bağımlı değişken Y, bağımsız değişken X ile ifade edilmiştir.

Tablo 4'te ayrıca EKK yöntemi, FCM ve G-K algoritmalarından elde edilen tahminler ve bu tahminlere ilişkin hata miktarları yer almaktadır.

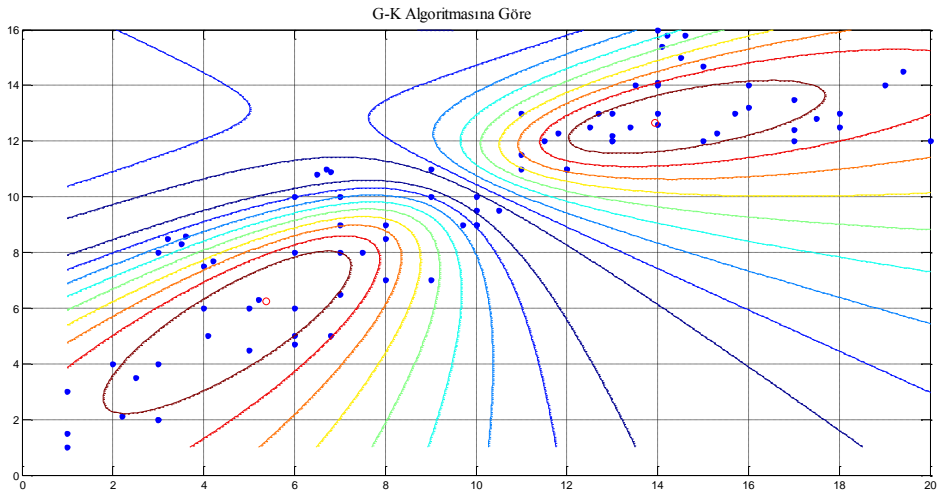
Tablo 4. Verilerin üç farklı yöntem ile tahminleri ve hataları

X	Y	\hat{Y}_{EKK}	e_{EKK_i}	\hat{Y}_{FCM}	e_{FCM_i}	\hat{Y}_{GK}	e_{GK_i}
1.0000	3.0000	4.0492	-1.0492	2.1098	0.8902	2.1098	0.8902
3.0000	2.0000	5.3492	-3.3492	4.0988	-2.0988	4.0988	-2.0988
2.0000	4.0000	4.6992	-0.6992	3.1043	0.8957	3.1043	0.8957
3.0000	4.0000	5.3492	-1.3492	4.0988	-0.0988	4.0988	-0.0988
4.0000	6.0000	5.9993	0.0007	5.0934	0.9066	5.0934	0.9066
6.0000	5.0000	7.2993	-2.2993	7.0824	-2.0824	7.0824	-2.0824
6.0000	8.0000	7.2993	0.7007	7.0824	0.9176	7.0824	0.9176
7.0000	9.0000	7.9493	1.0507	8.0770	0.9230	8.0770	0.9230
8.0000	7.0000	8.5993	-1.5993	9.0715	-2.0715	9.0715	-2.0715
9.0000	7.0000	9.2493	-2.2493	10.0660	-3.0660	10.0660	-3.0660
8.0000	9.0000	8.5993	0.4007	9.0715	-0.0715	9.0715	-0.0715
10.0000	9.0000	9.8994	-0.8994	11.0605	-2.0605	11.0605	-2.0605
9.0000	10.0000	9.2493	0.7507	10.0660	-0.0660	10.0660	-0.0660
16.0000	13.2000	13.7995	-0.5995	13.2329	-0.0329	13.2329	-0.0329
5.0000	4.5000	6.6493	-2.1493	6.0879	-1.5879	6.0879	-1.5879
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14.0000	13.0000	12.4994	0.5006	12.8203	0.1797	12.8203	0.1797
14.0000	12.6000	12.4994	0.1006	12.8203	-0.2203	12.8203	-0.2203
11.5000	12.0000	10.8744	1.1256	12.3046	-0.3046	12.3046	-0.3046
11.8000	12.3000	11.0694	1.2306	12.3665	-0.0665	12.3665	-0.0665
14.0000	14.1000	12.4994	1.6006	12.8203	1.2797	12.8203	1.2797
13.0000	12.2000	11.8494	0.3506	12.6140	-0.4140	12.6140	-0.4140
13.4000	12.5000	12.1094	0.3906	12.6965	-0.1965	12.6965	-0.1965
17.0000	12.4000	14.4495	-2.0495	13.4391	-1.0391	13.4391	-1.0391
17.5000	12.8000	14.7745	-1.9745	13.5423	-0.7423	13.5423	-0.7423
20.0000	12.0000	16.3995	-4.3995	14.0580	-2.0580	14.0580	-2.0580
18.0000	12.5000	15.0995	-2.5995	13.6454	-1.1454	13.6454	-1.1454
19.0000	14.0000	15.7495	-1.7495	13.8517	0.1483	13.8517	0.1483
19.4000	14.5000	16.0095	-1.5095	13.9342	0.5658	13.9342	0.5658
14.0000	16.0000	12.4994	3.5006	12.8203	3.1797	12.8203	3.1797
HATA		$\epsilon_{EKK} = 3.5039$		$\epsilon_{FCM} = 2.3696$		$\epsilon_{GK} = 5.7773$	

Şekil 14 ve 15'te sırasıyla bulanık kümeleme algoritmaları FCM ve G-K algoritmalarına göre kümelenmiş veriler yer almaktadır.



Şekil 14. FCM algoritmasına göre kümelenmiş veriler

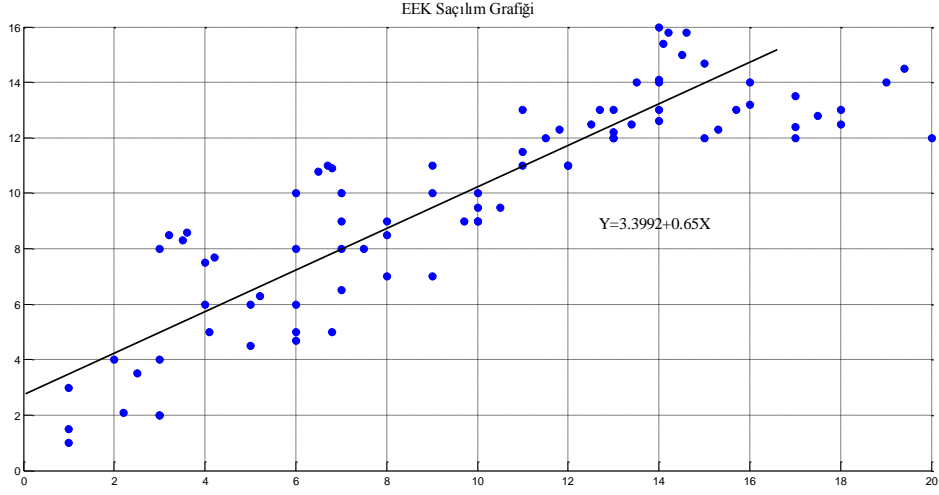


Şekil 15. G-K algoritmasına göre kümelenmiş veriler

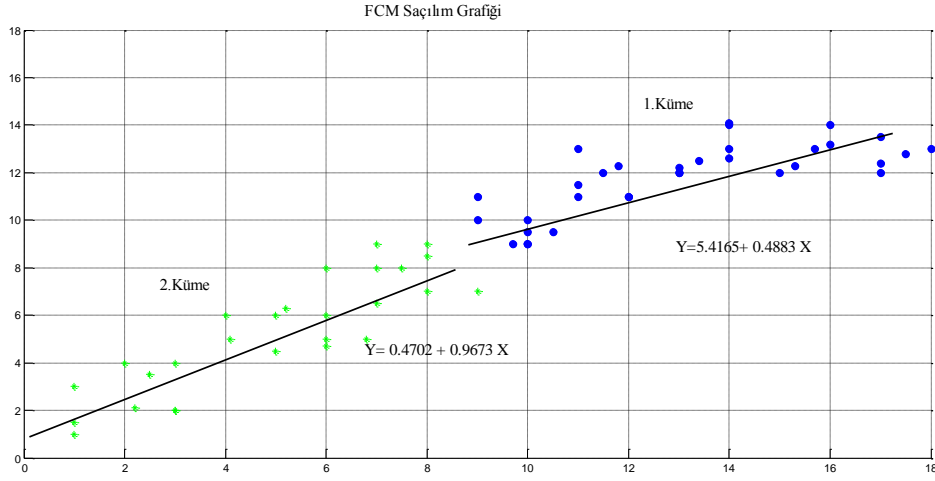
Şekil 16'da 85 verinin saçılım grafiğini ve En Küçük Kareler yöntemi ile elde edilen model yer almaktadır.

Şekil 17'de 85 veri Bulanık C-Ortalamlar yöntemi ile iki kümeye ayrılmış ve üyelik dereceleri kullanılması ile elde edilen modeller ayrı ayrı gösterilmiştir.

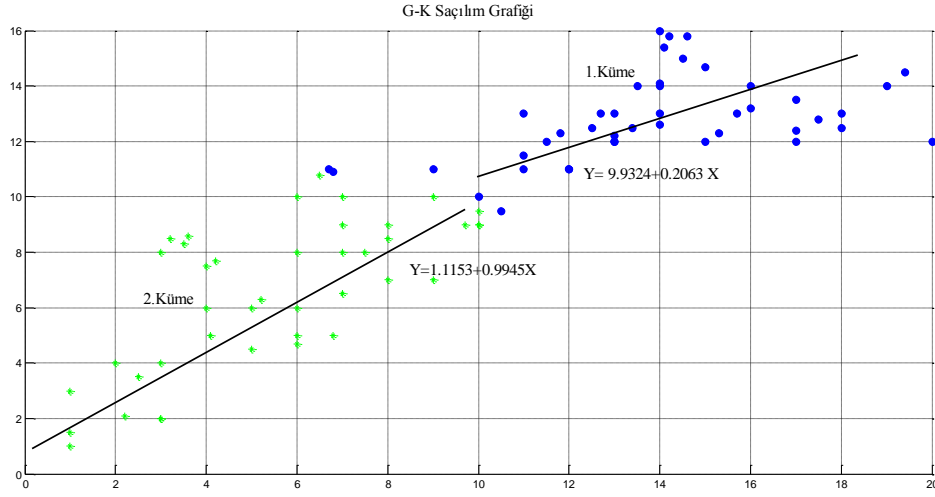
Şekil 18'de 85 veri Gustafson-Kessel yöntemi ile iki kümeye ayrılmış G-K algoritmasından elde edilen üyelik derecelerinin kullanılması ile oluşturulan modellere yer verilmiştir.



Şekil 16. EKK yöntemi ile elde edilen model



Şekil 17. FCM algoritması ile elde edilen modeller



Şekil 18. G-K algoritması ile elde edilen modeller

Tablo 5’te kullanılan üç yöntem ile elde edilen parametreler ile oluşturulan modeller yer almaktadır.

Tablo 5. Doğrusal regresyon modelleri ve hataları

En Küçük Kareler Yöntemi	Bulanık C-Ortalamalar	Gustafson-Kessel
$\hat{Y} = 3.3992 + 0.65 X$	$\hat{Y}_1 = 5.4165 + 0.4883X_{1j}$ $\hat{Y}_2 = 0.4702 + 0.9673X_{2j}$	$\hat{Y}_1 = 9.9324 + 0.2063X_{1j}$ $\hat{Y}_2 = 1.1153 + 0.9945X_{2j}$
$\varepsilon_{EKK} = 3.5039$	$\varepsilon_{FCM} = 2.3696$	$\varepsilon_{GK} = 5.7773$

Bulanık ağırlıklara dayalı parametre tahmini yaparken FCM ile kümelennmiş verilere ilişkin modelin hatası 2.3696, G-K ile kümelennmiş verilere ilişkin modelin hatası 5.7773 ve En Küçük Kareler yöntemi ile elde edilen modele ilişkin hata 3.5039 dir.

Bu örnekte de FCM ile elde edilen hata miktarı en küçük olarak gözlenmiştir.

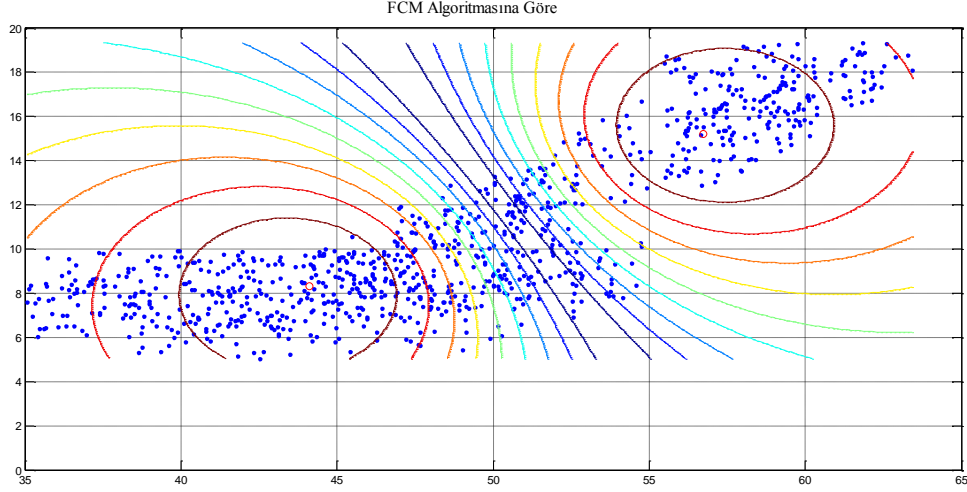
Örnek 3: Bu örnek iki aşamadan oluşmaktadır. İlk olarak gözlem setinde yer alan veriler için önerilen algoritma uygulanmış ve tahminler ve tahminlere ilişkin hata değerleri elde edilmiştir. İkinci aşamada ise algoritmanın aykırı değerlere olan duyarlılığını irdelleyebilmek için veri setine aykırı değerler eklenmiştir. Aykırı değerlerin de yer aldığı veri setine önerilen algoritma uygulanarak tahminler ve tahminlere ilişkin hata miktarları belirlenmiştir. Sözü edilen iki durum sırası ile Örnek 3.a ve Örnek 3.b de yer almaktadır. Örnek 3.a: Bu örnekte literatürde yer alan 944 gözlemden oluşan veri seti kullanılmıştır.

Verilerin bir kısmı Tablo 6’da yer almaktadır. Bir bağımlı ve bir bağımsız değişkene ilişkin gözlemlerden bağımlı değişken Y, bağımsız değişken X ile ifade edilmiştir. Tablo 6’da ayrıca EKK yöntemi, FCM ve G-K algoritmalarından elde edilen tahminler ve bu tahminlere ilişkin hata miktarları yer almaktadır.

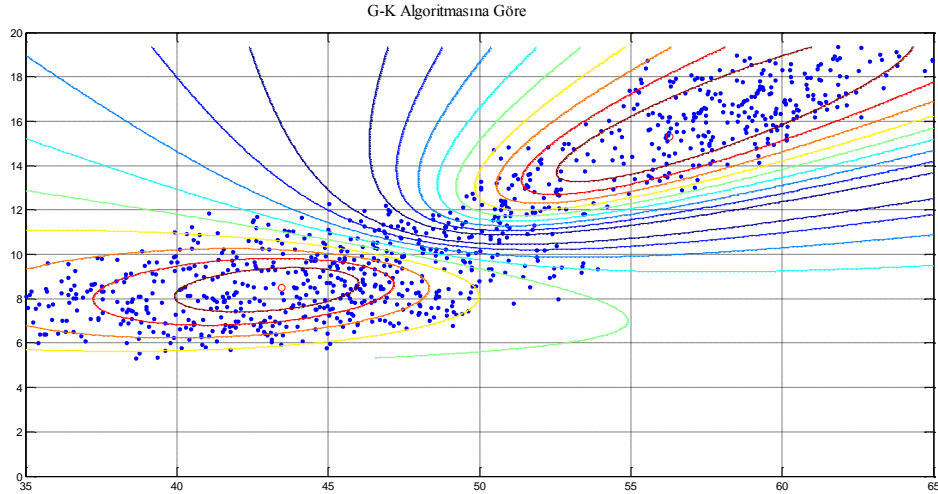
Tablo 6. Verilerin 3 Farklı Yöntem ile tahminleri ve hataları

X	Y	\hat{Y}_{EKK}	e_{EKK_i}	\hat{Y}_{FCM}	e_{FCM_i}	\hat{Y}_{GK}	e_{GK_i}
46.9800	11.8300	9.9224	1.9076	10.9745	0.8555	10.0268	1.8032
47.0500	11.9300	9.9540	1.9760	8.9607	2.9693	10.0665	1.8635
47.0700	11.2900	9.9631	1.3269	8.9652	2.3248	10.0779	1.2121
47.3700	10.7600	10.0985	0.6615	9.0336	1.7264	10.2482	0.5118
47.4000	11.5100	10.1121	1.3979	9.0404	2.4696	10.2652	1.2448
47.4100	10.7500	10.1166	0.6334	9.0457	1.7043	10.2709	0.4791
47.4600	11.1600	10.1392	1.0208	9.0541	2.1059	10.2993	0.8607
48.0100	10.6400	10.3875	0.2525	9.1793	1.4607	10.6114	0.0286
48.4600	11.5800	10.5907	0.9893	9.2818	2.2982	10.8669	0.7131
48.6500	11.7900	10.6765	1.1135	9.3251	2.4649	10.9747	0.8153
48.7600	10.4200	10.7262	-0.3062	9.3502	1.0698	11.0371	-0.6171
62.1150	18.0220	16.7570	1.2650	19.4885	-1.4665	18.6174	-0.5954
61.7150	18.9300	16.5763	2.3537	19.1664	-0.2364	18.3903	0.5397
61.5780	18.3040	16.5145	1.7895	19.0561	-0.7521	18.3126	-0.0086
61.2800	17.8440	16.3799	1.4641	18.8162	-0.9722	18.1434	-0.2994
60.7850	17.6650	16.1564	1.5086	18.4177	-0.7527	17.8625	-0.1975
60.3050	17.7760	15.9396	1.8364	18.0312	-0.2552	17.5900	0.1860
60.1010	16.7940	15.8475	0.9465	17.8670	-1.0730	17.4737	-0.6797
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
58.9680	16.6340	15.3359	1.2981	16.9548	-0.3208	16.8312	-0.1972
58.8190	15.7810	15.2686	0.5124	16.8348	-1.0538	16.7466	-0.9656
57.4690	15.9740	14.6590	1.3150	15.7480	0.2260	15.9804	-0.0064
56.3070	15.7000	14.1342	1.5658	14.8124	0.8876	15.3208	0.3792
54.3400	15.4830	13.2460	2.2370	13.2288	2.2542	14.2043	1.2787
52.5730	12.9040	12.4481	0.4559	11.8061	1.0979	13.2014	-0.2974
51.5420	13.1330	11.9825	1.1505	10.9761	2.1569	12.6162	0.5168
61.9020	18.8780	16.6608	2.2172	19.3170	-0.4390	18.4965	0.3815
35.4600	6.3900	4.7203	1.6697	7.5393	-1.1493	6.5811	-0.1911
35.6000	6.4100	4.7835	1.6265	6.3525	0.0575	6.6053	-0.1953
35.4000	6.0200	4.6932	1.3268	6.9721	-0.9521	6.5708	-0.5508
39.0400	7.3700	6.3369	1.0331	6.7967	0.5733	7.1986	0.1714
42.5090	8.1400	7.7142	0.4258	7.4550	0.6850	7.9263	0.2137
HATA		$\epsilon_{EKK} = 4.2197$		$\epsilon_{FCM} = 5.8365$		$\epsilon_{GK} = 3.1894$	

Şekil 19 ve 20’de sırasıyla bulanık kümeleme algoritmalarından FCM ve G-K algoritmalarına göre kümelenmiş veriler gösterilmektedir.



Şekil 19. FCM algoritmasına göre kümelenmiş veriler

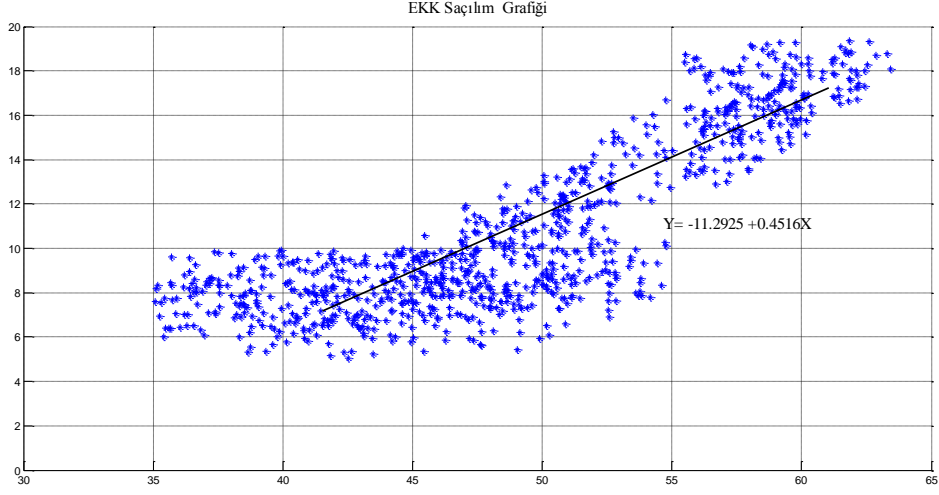


Şekil 20. G-K algoritmasına göre kümelenmiş veriler

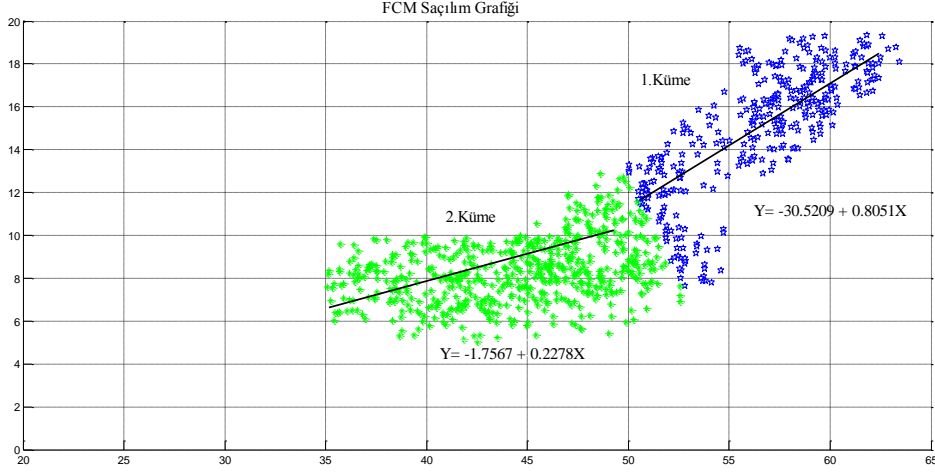
Şekil 21’de 944 verinin saçılım grafiğini ve En Küçük Kareler yöntemi ile elde edilen model yer almaktadır.

Şekil 22’de 944 veri Bulanık C-Ortalamalar yöntemi ile iki kümeye ayrılmış ve üyelik dereceleri kullanılması ile elde edilen modeller ayrı ayrı gösterilmiştir.

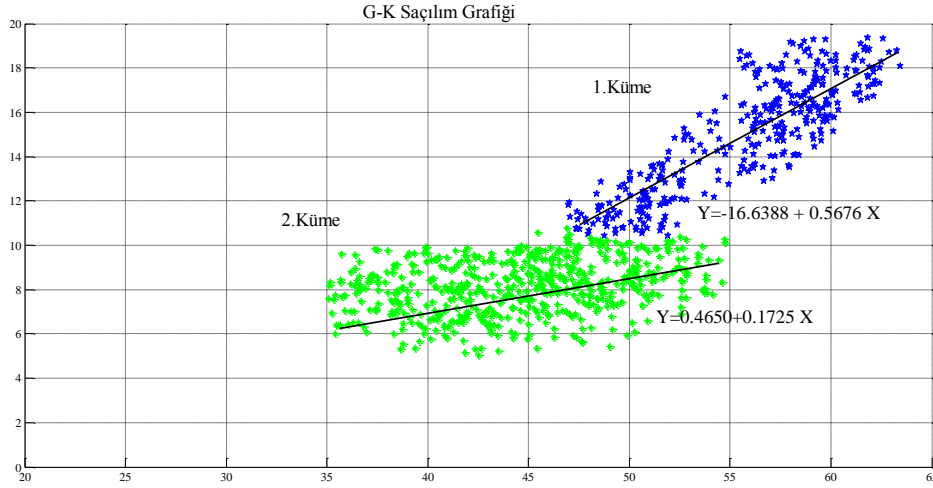
Şekil 23'te 944 veri Gustafson-Kessel yöntemi ile iki kümeye ayrılmış G-K algoritmasından elde edilen üyelik derecelerinin kullanılması ile oluşturulan modellere yer verilmiştir.



Şekil 21. EKK yöntemi ile elde edilen model



Şekil 22. FCM algoritması ile elde edilen modeller



Şekil 23. G-K algoritması ile elde edilen modeller

Tablo 7’de kullanılan üç yöntem ile elde edilen parametreler ile oluşturulan modeller yer almaktadır.

Tablo 7. Doğrusal regresyon modelleri ve hataları

En Küçük Kareler Yöntemi	Bulanık C-Ortalamalar	Gustafson-Kessel
$\hat{Y} = -11.2925 + 0.4516 X$	$\hat{Y}_1 = -30.5209 + 0.8051X_{1j}$ $\hat{Y}_2 = -1.7567 + 0.2278X_{2j}$	$\hat{Y}_1 = -16.638 + 0.5676X_{1j}$ $\hat{Y}_2 = 0.4650 + 0.1725X_{2j}$
$\epsilon_{EKK} = 4.2197$	$\epsilon_{FCM} = 5.8365$	$\epsilon_{GK} = 3.1894$

Bulanık ağırlıklara dayalı parametre tahmini yapılırken, FCM ile kümeleneş ve buradan elde edilen üyelik dereceleri kullanılarak ağırlıklandırılmış verilerden tahmin edilen parametreler ile kurulan modele ilişkin hata; eşitlik (54) ile verilen hata kriterlerinden 5.8365 olarak hesaplanmıştır. G-K ile kümeleneş ve buradan elde edilen üyelik dereceleri kullanılarak ağırlıklandırılmış verilerden tahmin edilen parametreler ile kurulan modele ilişkin hata; eşitlik (54) ile verilen hata kriterlerinden 3.1864 olarak hesaplanmıştır. EKK yöntemi ile elde edilen parametrelere ilişkin modelin hatası 4.2197 olarak belirlenmiştir.

Sonuç olarak G-K’ya dayalı modelden elde edilen tahminler en küçük hata miktarına sahiptir.

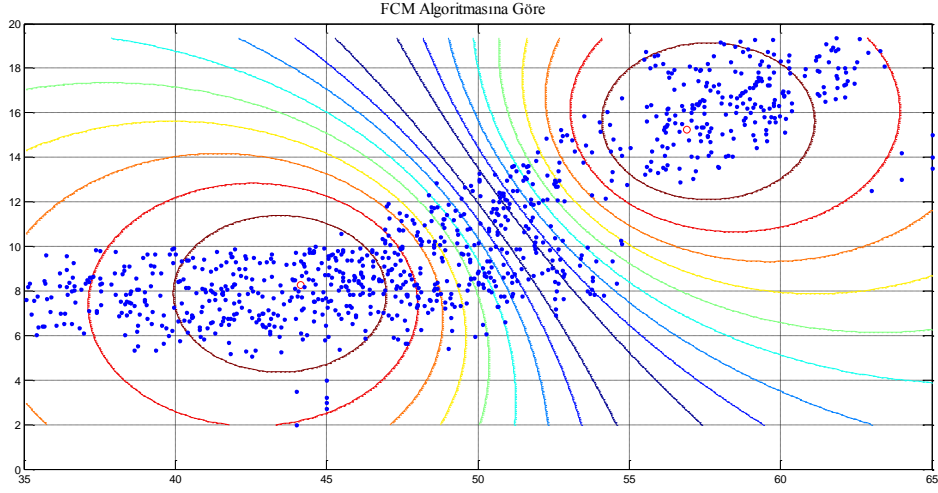
Örnek 3.b: Önerilen algoritmanın aykırı değere duyarlılığını gözlemlemek için 944 veriye 12 aykırı gözlem eklenmiştir.

Tablo 8’de EKK yöntemi, FCM ve G-K algoritmalarından elde edilen tahminler ve bu tahminlere ilişkin hata miktarları yer almaktadır.

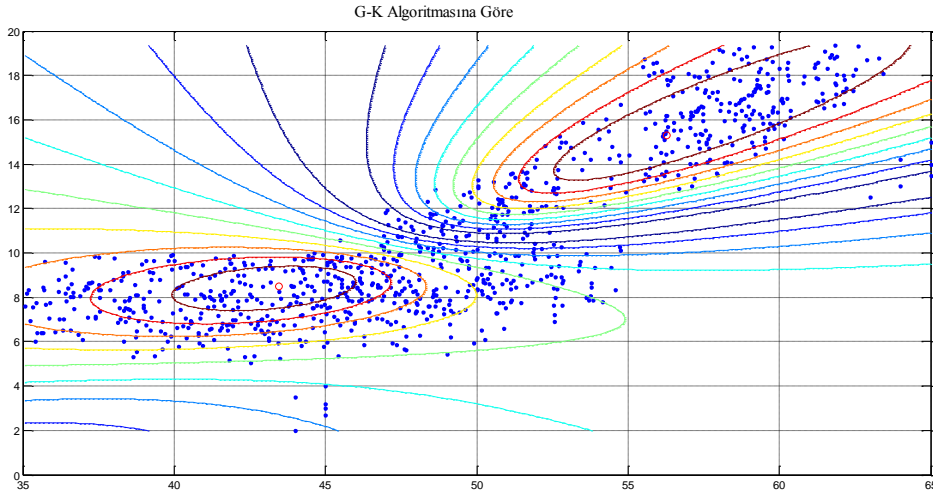
Tablo 8. Verilerin üç farklı yöntem ile elde edilen tahminleri ve hataları

X	Y	\hat{Y}_{EKK}	e_{EKK_i}	\hat{Y}_{FCM}	e_{FCM_i}	\hat{Y}_{GK}	e_{GK_i}
45.0000	3.2000	8.9847	-5.7847	8.4540	-5.2540	8.1825	-4.9825
45.0000	4.0000	8.9847	-4.9847	8.4540	-4.4540	8.1825	-4.1825
45.0000	2.7000	8.9847	-6.2847	8.4540	-5.7540	8.1825	-5.4825
44.0000	2.0000	8.5377	-6.5377	8.2120	-6.2120	8.0501	-6.0501
44.0000	3.5000	8.5377	-5.0377	8.2120	-4.7120	8.0501	-4.5501
45.0000	3.0000	8.9847	-5.9847	8.4540	-5.4540	8.1825	-5.1825
63.0000	12.5000	17.0300	-4.5300	19.8829	-7.3829	10.5643	1.9357
41.1830	8.2922	7.2787	1.0135	7.6774	0.6148	7.6774	0.6148
42.1030	9.0100	7.6899	1.3201	7.7991	1.2109	7.7991	1.2109
52.1400	10.2000	12.1760	-1.9760	9.1273	1.0727	9.1273	1.0727
51.9000	9.7300	12.0687	-2.3387	9.0955	0.6345	9.0955	0.6345
52.3100	8.9600	12.2520	-3.2920	9.1498	-0.1898	9.1498	-0.1898
52.4600	9.0500	12.3190	-3.2690	9.1696	-0.1196	9.1696	-0.1196
51.8900	9.8800	12.0643	-2.1843	9.0942	0.7858	9.0942	0.7858
51.2100	8.7500	11.7603	-3.0103	9.0042	-0.2542	9.0042	-0.2542
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50.3500	8.8500	11.3759	-2.5259	8.8904	-0.0404	8.8904	-0.0404
50.4100	8.6200	11.4028	-2.7828	8.8983	-0.2783	8.8983	-0.2783
49.6100	9.9900	11.0452	-1.0552	8.7925	1.1975	8.7925	1.1975
49.6900	8.3000	11.0810	-2.7810	8.8031	-0.5031	8.8031	-0.5031
48.8900	10.8400	10.7234	0.1166	10.7719	0.0681	10.7719	0.0681
49.4100	11.3700	10.9558	0.4142	11.0852	0.2848	11.0852	0.2848
49.1800	11.2100	10.8530	0.3570	10.9467	0.2633	10.9467	0.2633
42.3900	7.8900	7.8181	0.0719	7.8224	0.0676	7.8371	0.0529
50.9400	12.0800	11.6397	0.4403	12.0072	0.0728	12.0072	0.0728
51.0300	11.9100	11.6799	0.2301	12.0615	-0.1515	12.0615	-0.1515
52.6400	13.3800	12.3995	0.9805	13.0316	0.3484	13.0316	0.3484
61.5070	19.1040	16.3627	2.7413	18.3748	0.7292	18.3748	0.7292
60.8060	17.9730	16.0494	1.9236	17.9524	0.0206	17.9524	0.0206
58.6630	16.7640	15.0915	1.6725	16.6611	0.1029	16.6611	0.1029
HATA		$\epsilon_{EKK} = 4.4818$		$\epsilon_{FCM} = 6.9144$		$\epsilon_{GK} = 3.9836$	

Şekil 24 ve 25’te Bulanık kümeleme algoritmalarından FCM ve G-K algoritmalarına göre kümelenmiş veriler gösterilmektedir.



Şekil 24. FCM algoritmasına göre kümelenmiş veriler

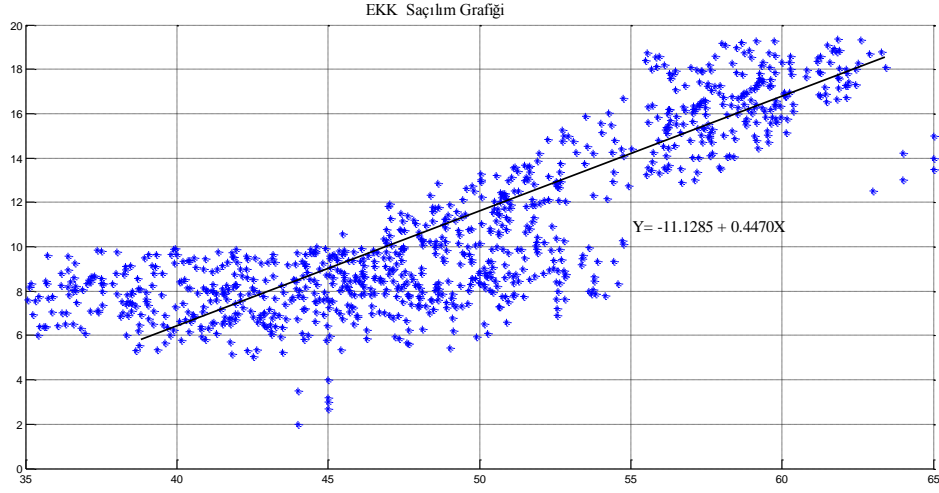


Şekil 25. G-K algoritmasına göre kümelenmiş veriler

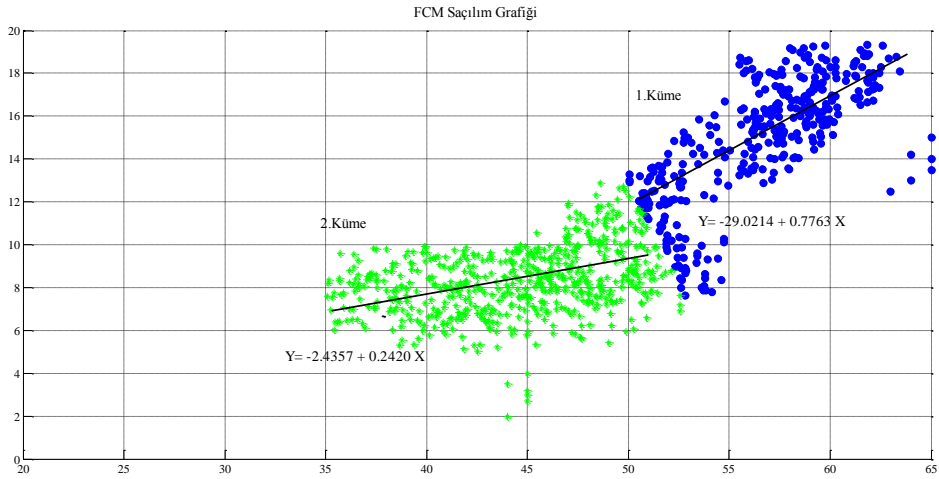
Şekil 26’de veri setinde yer alan verilerin saçılım grafiğini ve En Küçük Kareler Yöntemi ile elde edilen model yer almaktadır.

Şekil 27’de veri setinde yer alan verilerin Bulanık C-Ortalamalar yöntemi ile iki kümeye ayrılmış ve üyelik dereceleri kullanılması ile elde edilen modeller ayrı ayrı gösterilmiştir.

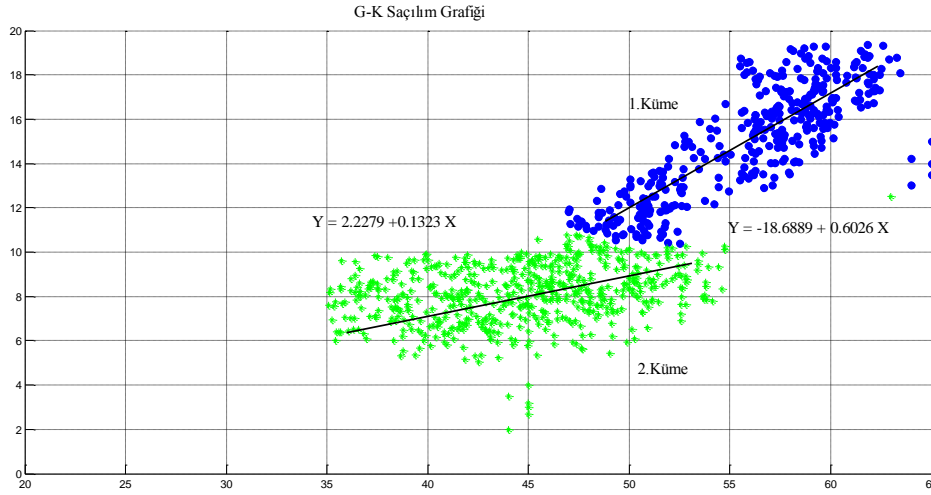
Şekil 28’de veri setinde yer alan veriler Gustafson-Kessel yöntemi ile iki kümeye ayrılmış G-K algoritmasından elde edilen üyelik derecelerinin kullanılması ile oluşturulan modellere yer verilmiştir.



Şekil 26. EKK yöntemi ile elde edilen model



Şekil 27. FCM algoritması ile elde edilen modeller



Şekil 28. G-K algoritması ile elde edilen modeller

Tablo 9’da kullanılan üç yöntem ile elde edilen parametreler ile oluşturulan modeller yer almaktadır.

Tablo 9. Doğrusal regresyon modelleri ve hataları

En Küçük Kareler Yöntemi	Bulanık C-Ortalamalar	Gustafson-Kessel
$\hat{Y} = -11.2985 + 0.4470X$	$\hat{Y}_1 = -29.0214 + 0.7763X_{1j}$ $\hat{Y}_2 = -2.4357 + 0.2420X_{2j}$	$\hat{Y}_1 = -18.6889 + 0.6026X_{1j}$ $\hat{Y}_2 = 2.2279 + 0.1323X_{2j}$
$\epsilon_{EKK} = 4.4818$	$\epsilon_{FCM} = 6.9144$	$\epsilon_{GK} = 3.9836$

Bu örnekte yer alan verilere ilişkin oluşturulan doğrusal regresyon modellerinden elde edilen tahminler ve hata kriterleri incelendiğinde, veri setine aykırı gözlem eklendiğinde de önerilen algoritmadan elde edilen sonuçların aykırı gözlemlerden EKK yöntemine nispeten daha az etkilendiği ve daha küçük hata miktarları ile tahminler yapabildiği gözlenmiştir.

4. SONUÇLAR VE ÖNERİLER

Üzerinde çalışılan verilerin aralarındaki fonksiyonel bağıntıyı modellemek için yapılacak parametre tahmini regresyon çözümlemesinin en önemli bölümlerinden biridir. Doğrusal regresyon modelinin, en küçük hataya sahip parametrelerinin elde edilmesi için pek çok yöntem önerilmiştir. Son yıllarda bilimin birçok alanında olduğu gibi regresyon çözümlemesinde de bulanık mantık yaklaşımları etkin bir biçimde kullanılarak parametre tahminleri yapılmaktadır.

Bu çalışmada veri setinde yer alan gözlemlerin iki kümeye ayrılabilmesi durumunda bölüm (1.6.2.2) de yer alan ve bulanık kümeleme yöntemlerine dayanan algoritmalar kullanılarak, parametre tahminleri yapılmış ve bu tahminlere ilişkin hata miktarları modellerin geçerliliklerinin sınanmasında gösterge olmuştur. Yapılan uygulamalardan da izlenebileceği gibi veri setinin tek bir kümede yoğunlaşmaması durumunda bulanık kümeleme yöntemlerinden elde edilen üyelik dereceleri ile ağırlıklandırılan bağımsız değişkenlerden doğrusal modeller elde edilmiştir. Bu modellere ilişkin hata miktarları EKK yönteminden elde edilen modele ilişkin hatalarla karşılaştırılmış ve önerilen algoritmanın etkinliği verilerin tek bir kümede yoğunlaşmaması ve veri setinde aykırı gözlemlerin varlığı durumunda parametre tahmininde alternatif yöntem olarak kullanılabilmesi gösterilmiştir.

Yöntem, bağımsız değişken sayısının iki ve daha fazla olması ve değişkenlerin toplandığı küme sayılarının ikiden çok olması durumları için genelleştirilerek etkinlikleri incelenebilir.

5. KAYNAKLAR

- Alpar, R., 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, 2. Baskı, Nobel Yayınevi, Ankara.
- Apaydın, A., Kutsal, A. ve Atakan, C., 1994. Uygulamalı İstatistik, 496s., Ankara.
- Apaydın, A., Hancı, B., Y. ve Dalkılıç Erbay, T., 2010. Fuzzy adaptive neural network approach to path loss prediction in urban areas at GSM-900 band, Turk J Elec Eng & Comp Sci,18,6,1077-1094.
- Atbaş., G., A., C., 2008. Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Avcı, U., 2006. Bulanık Kümeleme Algoritmalarının Karşılaştırılmalı Analizi ve Bilgisayar Uygulamaları, Yüksek Lisans Tezi, Ege Üniversitesi, Fen Bilimleri Enstitüsü, Bornova-İzmir.
- Ay., M., 2010. Sulama Suyu Kalitesinin Kümelemeye Dayalı Bulanık Sistem ile Sınıflandırılması, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Fen Bilimleri Enstitüsü, Sivas.
- Başer, F., 2007. Aktüeryal Modellemede Melez Bulanık Regresyon Analizi, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Baykal, N. ve Beyan, T., 2004. Bulanık Mantık İlke ve Temelleri, Bıçaklar Kitabevi, Ankara.
- Bezdek, C., J., 1980. A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms, IEEE Trans. Pattern Anal. Machine Intell. Vol. PAMI-2,1-8.
- Bezdek, J., C., Ehrlich, R. ve Full, W., 1984. Fcm: The Fuzzy c-Means Clustering Algorithm, Computers & Geosciences,10,191-203.
- Büyükköz, D., 2010. Fuzzy Kümeleme Teknikleri ve Avrupa Birliği Üye Ülkeleri ile Türkiye'nin Fuzzy Kümelenmesi, Yüksek Lisans Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Cheng, C. and Lee, E., S., 2001. Switching regression analysis by fuzzy adaptive network, European Journal of Operational Research,128, 647-663.
- Dombi, J., 1990. Membership Functions As An Evaluation, Fuzzy Sets and Systems, Vol 35,1-21
- Düzyurt, S., 2008. Bulanık Regresyon ile Tahmin ve Bir Uygulama, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Fakültesi, Ankara.

- Eını., S., 2009. Bulanık Doğrusal Regresyon Uygulaması, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Elmas, Ç., 2003. Bulanık Mantık Denetleyiciler, Seçkin Yayınevi, 22-40, 230s.,Ankara.
- Erbay Dalkılıç., T., 2005. Switching Regresyon’da Bulanık Sinir Ağları Yaklaşımı ile Parametre Tahmini, Doktora Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Erbay, T., 1999. Minmad Regresyon Problemine Sinir Ağları Yaklaşımı, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Ergün, E., 2011. Bulanık Regresyon ve Yarıparametrik Toplamsal Regresyon Üzerine Bir Örnek Uygulama Çalışması, Yüksek Lisans Tezi, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Muğla.
- Erilli, N., A., 2009. Kümeleme Analizine Bulanık Yaklaşım Algoritmaları ve Uygulamaları, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Samsun.
- Erilli, N., A., Uslu, V., R, Yolcu, U., Aladağ, C., H. ve Eğrioğlu, E. 2011. Fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering, Expert Systems With Applications 38,10355-10357.
- Ertan, E., 2008. Regresyon Analizi ve Matematik Programlama Arasındaki İlişki, Yüksek Lisans Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Gath, L. and Geva, A., B., 1989. Unsupervised Optimal Fuzzy Clustering, IEEE Transactions on Pattern Analysis Machine Intelligent,11,7,773-781.
- Gustafson, D., E. and Kessel, W., C., 1979. Fuzzy Clustering With A Fuzzy Covariance Matrix, IEEE Conf. Decision Contr.,761-766.
- Güler, N., 2006. Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları, Yüksek Lisans Tezi, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Muğla.
- Gün., A., M., 2011. Bulanık Kümeleme Analiziyle Görüntüdeki Yüzün Tanınması, Yüksek Lisans Tezi, Haliç Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Günay, Atbaş, A., C., 2008. Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- İşbilen, Yücel, L., 2005. Bulanık Regresyon: Türkiye’de 1980-2004 Döneminde Kayıt Dışı Ekonominin Bulanık Yöntemlerle Tahminine İlişkin Uygulama, Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Kaya, S., H., 2010. Bulanık Regresyon ve Bir Uygulama, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü, Isparta.

- Kocaturk., Y., 2007. Bulanık Değişkenler ve Bulanık Yenileme Süreçleri, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Krishnapuram, R. and Kim, J., 1999. A Note on the Gustafson-Kessel and Adaptive Fuzzy Clustering Algorithms, IEEE Transactions on Fuzzy Systems,7, 453-461.
- Kruse, R. and Lesot, M., J., 2008. Gustafson-Kessel-like clustering algorithm based on typicality degree, Uncertainty and Intelligent Information System,117- 130.
- Özkan, M., M., 2003. Bulanık Hedef Programlama, Ekin Kitabevi, Bursa.
- Pham, T., D., 2011. Clustering Data with Spatial Continuity, 2001 IEEE International Fuzzy Systems Conference, 69-72.
- Saraçoğlu, R., 2007. Bulanık Kümeleme Kullanarak Benzer Belge Araması, Doktora Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- Şanlı, K., 2005. Bulanık Robust Regresyon Çözümlemesi, Doktora Tezi, Ankara Üniversitesi, Ankara.
- Şen, Z., 2001. Bulanık Mantık ve Modelleme İlkeleri, Bilge Sanat Yapım Evi, İstanbul.
- Turusbekova., A., 2012. Bulanık Kümeleme Algoritmalarına Dayalı Mamografik Kitle Sınıflandırma, Yüksek Lisans Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Ulutağay, G., 2004. Bulanık C-Ortalamlar Kümeleme Analizi ve Uygulamaları, Yüksek Lisans Tezi, Ege Üniversitesi, Fen Bilimleri Enstitüsü, Bornova-İzmir.
- Ünver, Ö., Gamgam., H. ve Altunkaynak., B., 2011. Temel İstatistik Yöntemler, 6.Baskı, 424s., Seçkin Kitapevi, Ankara.
- Windham., M., P., 1983. Geometrical Fuzzy Clustering Algorithms, Reprinted with Permission from Fuzzy Sets and Systems,10,271-279.
- Xie, X., L. and Beni, G.,1991. A Validity Measure for Fuzzy Clustering, IEEE Transactions On Pattern Analysis and Machine Intelligence,13,8,841-847.
- Yapıcı., N., 2000. Bulanık Doğrusal Programlamaya Sınır Ağları Yaklaşımı, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- Zadeh, L., A.,1965. Fuzzy Sets, Information and Control,8,338-353.
- Zahid, N., Limouri M. and Essaid, A., 1999. A New cluster-validity for fuzzy clustering, Pattern Recognition,32,1089-1097.
- Zehl, K., Einax J., W. and Sarbu, C., 2007. Fuzzy divisive hierarchical clustering of soil data Using Gustafson-Kessel algorithm, Chemometrics and Intelligent Laboratory System,86, 121-129.

ÖZGEÇMİŞ

Ersegül GÖKTÜRK ŞAHİN, 09 Temmuz 1986 tarihinde Trabzon'da doğdu. İlköğrenimini 24 Şubat İlköğretim Okulu'nda, ortaöğrenimini Cumhuriyet Orta Okulu'nda, lise öğrenimini Tevfik Serdar Anadolu Lisesi'nde tamamladı.2005-2006 öğretim yılında Karadeniz Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü'ne yerleşti. 2009 yılında mezun olduktan sonra aynı yıl aynı bölümde tezli yüksek lisans programına dahil oldu. 2010 yılında Türkiye İş Bankasında çalışmaya başladı ve halen bu görevine devam etmektedir.